Semiparametric Robust Estimation of Truncated and Censored Regression Models

Cizek, P.

Publication date:
2008

Citation for published version (APA):
Cizek, P. (2008). *Semiparametric Robust Estimation of Truncated and Censored Regression Models*. (CentER Discussion Paper; Vol. 2008-34). Econometrics.

CentER

Discussion Paper

No. 2008–34

SEMIPARAMETRIC ROBUST ESTIMATION OF TRUNCATED AND CENSORED REGRESSION MODELS

By Pavel Čížek

March 2008

ISSN 0924-7815

TILBURG UNIVERSITY

Semiparametric Robust Estimation of Truncated and Censored Regression Models

Pavel Čížek[1]

Department of Econometrics and Operation Research

Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands[1]

Abstract

Many estimation methods of truncated and censored regression models such as the maximum likelihood and symmetrically censored least squares (SCLS) are sensitive to outliers and data contamination as we document. Therefore, we propose a semiparametric general trimmed estimator (GTE) of truncated and censored regression, which is highly robust and relatively imprecise. To improve its performance, we also propose data-adaptive and one-step trimmed estimators. We derive the robust and asymptotic properties of all proposed estimators and show that the one-step estimators (e.g., one-step SCLS) are as robust as GTE and are asymptotically equivalent to the original estimator (e.g., SCLS). The finite-sample properties of existing and proposed estimators are studied by means of Monte Carlo simulations.

Keywords: asymptotic normality, censored regression, one-step estimation, robust estimation, trimming, truncated regression

JEL codes: C13, C14, C21, C24

1 Introduction

In statistics and econometrics, more attention has been recently paid to techniques that can deal with data contamination and outliers, which can arise from miscoding or heterogeneity not captured or presumed in a model. Evidence for outliers and data contamination of a part of data and its adverse effects on estimators based on the least squares (LS) or maximum likelihood (MLE) principles is provided, for example, by Gerfin (1996) in labor market

[1] *Corresponding author.* Tel.: +31 13 466 8723. Fax: +31 13 466 3280. Email: P.Cizek@uvt.nl.

data, by Peracchi (1990) in household income-expenditure data, and by Sakata and White (1998) in financial time series. The need for estimation procedures insensitive to data contamination and large errors have been recognized by many authors, for example, Krasker and Welsch (1985), Hampel et al. (1986), Peracchi (1990), Krishnakumar and Ronchetti (1997), Ronchetti and Trojani (2001), Ortelli and Trojani (2005), and Bramati and Croux (2007). In this paper, we address robust estimation of truncated and censored regression models. On one hand, we document the sensitivity of existing (semi)parametric estimators to outliers and data contamination. On the other hand, we propose new highly robust semiparametric estimators of truncated and censored regression, derive their robust and asymptotic properties, and document in simulations that the proposed estimators provide robust and stable results without sacrificing the finite-sample performance.

The classical MLE estimation of truncated and censored regression is sensitive to departures from the assumptions of normality and homoscedasticity (Arabmazar and Schmidt, 1981). This gave rise to semiparametric estimators based on weaker identification assumptions such as symmetrically trimmed least squares (STLS) and symmetrically censored least squares (SCLS), which rely on the conditional symmetry of errors (Powell, 1986), censored least absolute deviations (CLAD) estimator, which assumes the conditional median of errors being zero (Powell, 1984; Khan and Powell, 2001), and the mode regression (Lee, 1993). These concepts were later extended to the panel-data context (Honore, 1992; Honore and Powell, 1994). Further extensions include quantile regression (Portnoy, 2003), random censoring (Honore et al., 2002), and models with endogeneity (Hu, 2002; Honore and Hu, 2007). Alternative methods based on nonparametric estimation of the density function include those by Gallant and Nychka (1989), Ichimura (1993), Lewbel (1998), and Lewbel and Linton (2002).

Many truncated and censored regression estimators such as STLS or CLAD are often regarded to be robust not only in terms of identification assumptions, but also to outliers and data contamination because they employ "trimming" of regression residuals. This however holds only to a limited extent (Peracchi, 1990; Santos Silva, 2001) as we also document in this paper. This gave rise to a number of robust estimators of truncated and censored regression. For example, Peracchi (1990), Zhou (1992), Ren and Gu (1997), and Ren (2003) proposed various robust M-estimators for censored data, which bound the MLE score function to achieve lower sensitivity to extreme observations. Adjusting MLE however imposes strong identification assumptions as MLE itself and the methods cannot be applied under random regressors and heteroscedasticity, for instance. Additionally, the robustness of M-estimators is typically very limited as the number of explanatory variables increases (Maronna et al., 1979) unless model-independent trimming of observations is used. Although the second concern could be eliminated by using a robust truncated MLE (Čížek, 2007b; Marazzi and Yohai, 2004), the strong identification assumptions inherent to MLE are still present. Therefore, Debruyne et al. (2008) applied the concept of regression depth (Rousseeuw and Hubert, 1999) to censored quantile regression to create a robust alternative of CLAD. This method is however of a limited use given that there is no asymptotic theory, linear-regression depth was studied only in the i.i.d. case, and even a reliable computational algorithm does not exist.

In this paper, we propose robust estimators of truncated and censored regression, which are based on the STLS and SCLS estimators in order to construct semiparametric robust methods under weak identification assumptions (although the proposed concept can be applied also to CLAD and other semiparametric estimators). In the truncated-regression case, we start from STLS, which symmetrically trims regression residuals, and generalize it to "trimmed STLS" by including an additional kind of trimming, which protects against outlying observations (a data-adaptive choice of the trimming amount is proposed as well). As a by-product, a robust estimator of the residual variance in the truncated and censored regression is developed. In the censored-regression case, this approach is not applicable, and therefore, we propose the one-step SCLS, which performs one step of the SCLS computational algorithm starting from an initial robust estimator, for example, the trimmed STLS. Performing just one step preserves robust properties of the initial estimator, whereas using the SCLS iterative formula allows to employ information from all sample observations. Next, we study both robust and asymptotic properties of all proposed estimators, and in particular, we show that the data-adaptive and one-step estimators are asymptotically equivalent to the original STLS and SCLS if there are no outlying observations. Although we mostly restrict ourselves to cross-sectional models, the proposed estimation methods can be straightforwardly generalized to CLAD, panel data, and other models considered in the above discussed literature on extensions of STLS, SCLS, and CLAD estimators.

The paper is organized in the following way. We first introduce some existing estimators of truncated and censored regression and basic concepts regarding the robust estimation in

Section 2. The proposed robust estimators are introduced in Section 3, where we also study their robust properties. The asymptotic distributions of all proposed methods are derived in Section 4. Finally, all estimation methods are compared in finite samples by means of Monte Carlo simulation in Section 5. The proofs are provided in the Appendix.

2 Estimation of truncated and censored regression

Let us now introduce various parametric and semiparametric estimators of truncated and censored regression and discuss their robust properties (Sections 2.1). Later, we introduce the concept of the general trimmed estimator (GTE), which will render the robust alternatives to some well-known estimators of truncated and censored regression (Sections 2.2).

2.1 Truncated and censored regression

We consider the latent linear regression model

$$y_i^* = x_i^\top \beta^0 + \varepsilon_i, \tag{1}$$

where $y_i^* \in \mathbb{R}$ is the latent (unobservable) response variable, $x_i \in \mathbb{R}^p$ represents a vector of explanatory variables (including intercept), β^0 denotes the true value of the parameter vector $\beta \in \mathbb{R}^p$, and ε_i is the latent error term with standard deviation $\sigma = \sqrt{\mathsf{var}(\varepsilon_i)}$. Without loss of generality, we assume that the truncation or censoring occurs at zero from below. In the case of truncation, this means that the we observe only data points $(y_i, x_i) = (y_i^*, x_i)$ such that $y_i^* > 0$; this truncated model will be denoted by TM. In the case of censoring, we observe data points (y_i, x_i) with response $y_i = \max\{y_i^*, 0\}$; the resulting model $y_i = \max\{x_i^\top \beta + \varepsilon_i, 0\}$ will be referred to as CM.

Denoting $d_i = I(y_i > 0)$, the parameter vector β^0 can be estimated by MLE, which maximizes in the truncated-regression case (model TM)

$$\ln L_n(\beta, \sigma) = \sum_{i=1}^{n} \left\{ \ln f_\sigma^*(y_i - x_i^\top \beta) - [1 - \ln F_\sigma^*(0 - x_i^\top \beta)] \right\} \tag{2}$$

and in the censored-regression case (model CM)

$$\ln L_n(\beta, \sigma) = \sum_{i=1}^{n} \left\{ d_i \ln f_\sigma^*(y_i - x_i^\top \beta) + (1 - d_i) \ln F_\sigma^*(0 - x_i^\top \beta) \right\}, \tag{3}$$

where we typically assume $f_\sigma^*(t) = \phi(t/\sigma)$ and $F_\sigma^*(t) = \Phi(t/\sigma)$ (ϕ and Φ represent the standard normal density and distribution functions, respectively).

Since this MLE estimation is extremely sensitive to the violation of distributional assumptions (Arabmazar and Schmidt, 1982), many alternative semiparametric estimators have been proposed. In the case of truncation, Powell (1986) proposed the STLS estimator minimizing

$$\frac{1}{n} \sum_{i=1}^{n} [y_i - \max(y_i/2, x_i^\top \beta)]^2, \tag{4}$$

which relies on the conditional symmetry of the ε_i distribution. The same assumption in the censored regression model leads to the SCLS estimator minimizing (Powell, 1986)

$$\frac{1}{n} \sum_{i=1}^{n} \left\{ [y_i - \max(y_i/2, x_i^\top \beta)]^2 + I(y_i > 2x_i^\top \beta)[(y_i/2)^2 - \max(0, x_i^\top \beta)^2] \right\}. \tag{5}$$

Other alternatives in the context of censored regression are, for example, the CLAD estimator (Powell, 1984), which is based on the conditional median restriction, and various semiparametric estimators based on nonparametric estimation of the density function (Gallant and Nychka, 1989; Lewbel and Linton, 2002). In this paper, we concentrate on the STLS and SCLS estimators, but many results are valid for or can be generalized to many other estimators of truncated and censored regression models such as CLAD or extensions of STLS and SCLS (e.g., Honore, 1992; Honore and Powell, 1994).

The discussed semiparametric estimators, STLS and SCLS, are often considered to be robust to data contamination because the symmetric trimming places an upper bound on the contribution of each observation to the objective function. In the case of STLS, for example, the contribution of an observation (y_i, x_i) cannot exceed $(y_i/2)^2$, see (4). We will however document that all introduced estimators can be arbitrarily biased (towards zero or infinity) by data contamination anyway.

To formulate a result concerning the global robustness of truncated- and censored-regression estimators, we have to introduce a formal definition of the breakdown point, which measures the smallest fraction of observations that, added at appropriate locations, can make the estimator "useless." For the sake of simplicity, we consider independent and identically distributed observations $(y_i, x_i)_{i=1}^n$ (the breakdown point under dependence is generally model-specific; see Genton and Lucas, 2003). The finite-sample breakdown point of a truncated or censored regression estimator $\hat{\beta}_n(Z_n)$ at sample $Z_n = (x_i, y_i)_{i=1}^n$ can be then defined as (Rousseeuw, 1997)

$$\epsilon_n^* = \frac{1}{n} \min\left\{ m \in \mathbb{N}_0 : \sup_{Z_n'} \|\hat{\beta}_n(Z_n')\| = \infty \right\}, \tag{6}$$

where Z_n' represents samples obtained from Z_n by replacing any m observations by arbitrary values. The asymptotic breakdown point of the estimator $\hat{\beta}_n$ is then the corresponding limit $\epsilon^* = \lim_{n\to\infty} \epsilon_n^*$ (it usually exists and is independent of the data-generating process for cross-sectional data). The breakdown point of a scale estimator $\hat{\sigma}_n(Z_n)$ can be defined similarly:

$$\epsilon_n^* = \frac{1}{n} \min\left\{ m \in \mathbb{N}_0 : \inf_{Z_n'} \|\hat{\sigma}_n(Z_n')\| = 0 \text{ or } \sup_{Z_n'} \|\hat{\sigma}_n(Z_n')\| = \infty \right\}.$$

Now, we show that the breakdown points of MLE, STLS, and SCLS in the truncated and censored regression models are asymptotically equal to zero, especially if data contamination includes leverage points, that is, observations with large values of explanatory variables (this result can be proved in the same way also for CLAD and other censored-regression estimators). Additionally, let us note that proving $\sup_{Z_n'} \|\hat{\beta}_n(Z_n')\| = \infty$ in (6) does not necessarily imply that the slope estimates have to diverge: in the proof of the following theorem, we construct samples Z_n' such that all slope estimates become arbitrarily close to zero.

Theorem 1 *Let $(x_i, y_i)_{i=1}^n$ be a sequence of independent and identically distributed random vectors from the truncated-regression model TM or censored-regression model CM. We assume that the data are almost surely in a general position for $n > 2p$ and that the models include intercept. Then the finite-sample breakdown points of the MLE and STLS estimators in truncated regression and of the MLE and SCLS estimators in censored regression are smaller than or equal to $\epsilon_n^* = 2p/n$, which tends to $\epsilon^* = 0$ as $n \to \infty$.*

2.2 General trimmed estimator

The parametric and semiparametric estimators of the truncated and censored regression models TM and CM are sensitive to outlying observations, especially those with high leverage (see the proof of Theorem 1). One of traditional solutions reducing or eliminating this sensitivity in (non)linear regression amounts to downweighting or trimming observations that have large regression residual (e.g., Rousseeuw, 1985). We will introduce here a generalization of this concept – the general trimmed estimation – to facilitate the proposal of semiparametric robust regression estimators later in Section 3.

To address high sensitivity of the MLE- and LS-based methods to outlying or misspecified observations, Čížek (2007b) proposed the concept of general trimmed estimation (GTE). Given a sample $(x_i, y_i)_{i=1}^n$ and an estimation method T that minimizes the objective function of the form $\sum_{i=1}^n s(x_i, y_i; \beta)$ over $\beta \in B$, where $s(x_i, y_i; \beta)$ represents a loss function identifying the true value β^0 of parameter vector β, one typically knows that small values of $s(x_i, y_i; \beta)$ represent likely observations under a given model ("good fit," e.g., small squared residuals) and large values of $s(x_i, y_i; \beta)$ correspond to unlikely values ("bad fit," e.g., large squared residuals). For example, if we observe y_i^* and x_i in the latent linear model (1), we could estimate it by LS using $s(x_i, y_i^*; \beta) = (y_i^* - x_i^\top \beta)^2$; the small residuals would then mean that a given observation is fit by the model well and vice versa.

To create a method insensitive to outliers and observations badly explained by the model, GTE minimizes an objective function from which the unlikely observations (i.e., observations with large values of $s(x_i, y_i; \beta)$), are trimmed away. In this simple case, the general trimmed estimator $\hat{\beta}_n^{(GTE\text{-}T,h)}$ obtained from the estimation method T can therefore be defined as

$$\hat{\beta}_n^{(GTE\text{-}T,h)} = \arg\min_{\beta \in B} \sum_{j=1}^{h_n} s_{[j]}(\beta), \tag{7}$$

where $s_{[j]}(\beta)$ represents the jth smallest order statistics of $s(x_i, y_i; \beta)$, $i = 1, \ldots, n$. Thus, the GTE estimate minimizes the loss of h_n most likely observations under a given parametric model, where the trimming constant satisfies $n/2 < h_n \leq n$. Note that the trimming constant h_n determines the insensitivity to outliers and breakdown point of GTE because (7) indicates that $n - h_n$ observations with the largest losses do not directly affect the estimator. In the (non)linear regression, GTE combined with LS (GTE-LS) results in the well-known least trimmed squares estimator (LTS; Rousseeuw, 1985), which achieves the maximum asymptotic breakdown point 1/2 if $h_n = [(n+1)/2] + p$ (Rousseeuw and Leroy, 2003; Stromberg and Ruppert, 1992); $[z]$ denotes the integer part of z.

3 Robust estimation of truncated and censored regression

In Section 2, we have seen that many well-known estimators of truncated and censored regression are sensitive to outliers and data contamination and have asymptotically breakdown points equal to zero (Theorem 1). At the same time, a general GTE concept of creating robust estimators was introduced. Since MLE relies on strong identification assumptions, we start by applying GTE to the less-restrictive STLS estimator in truncated regression (Section 3.1). Because trimming of observations usually results in a substantial increase of the variance of estimates, we also propose a data-adaptive procedure for the choice of trimming amount to eliminate or minimize the relative-efficiency loss (Section 3.2). Finally, we discuss how the proposed robust estimation methods can be extended to the SCLS estimator in the censored regression model (Section 3.3). The identification assumptions and asymptotic properties of all proposed methods will be studied later in Section 4.

3.1 Robust estimation of truncated regression models

To propose a robust alternative to the MLE and STLS estimators of the truncated-regression model TM, we start from STLS, which relies on relatively weak identification assumptions, and apply the GTE concept to STLS. More precisely, we propose the GTE-STLS estimator of the model TM defined by

$$\hat{\beta}_n^{(GTE\text{-}STLS,h_n)} = \arg\min_{\beta \in B} \sum_{j=1}^{h_n} s_{[j]}(\beta) = \arg\min_{\beta \in B} \sum_{j=1}^{h_n} \left\{ [y_i - \max(y_i/2, x_i^\top \beta)]^2 \right\}_{[j]} \quad (8)$$

where $s_{[j]}(\beta)$ represents the jth smallest order statistics of $s(x_i, y_i; \beta) = [y_i - \max(y_i/2, x_i^\top \beta)]^2$, $i = 1, \ldots, n$. (Note that we could obviously define the same way a GTE counterpart of MLE.) An important feature of GTE-STLS is that it trims the observations with large absolute values of symmetrically trimmed residuals $\min\{y_i/2, y_i - x_i^\top \beta\}$ rather than only the observations with observable residuals, $0 \leq x_i^\top \beta$. Similarly to LTS, the proposed GTE-STLS trims exactly $n - h_n$ observations from the objective function, where $[(n+1)/2] + p \leq h_n \leq n$, and can thus survive contamination of up to $[(n-h_n)/n]$ percent of data (cf. Stromberg and Rupert, 1992), where the most robust choice of trimming is $h_n = [(n+1)/2] + p$ and $[(n-h_n)/n] \approx 0.5$.

3.2 Adaptive choice of trimming

The GTE-STLS estimator proposed in Section 3.1 is a robust alternative to MLE and STLS. It is however well known that trimming 25% or 50% observations of the sample causes a substantial increase in the variance of estimates even in the linear regression (cf. Čížek, 2007a) and we can thus expect en even more negative effect of trimming on the relative efficiency of the estimator in the truncated regression case. Therefore, we complement GTE-STLS by a data-adaptive choice of the trimming constant h_n so that the smallest possible amount $n - h_n$ of observations is trimmed from the objective function.

The adaptive choice of trimming is motivated by Gervini and Yohai (2002) and Čížek (2007a), who studied the data-dependent trimming in the context of the linear regression model. It is therefore beneficial to describe the linear-regression case first. Provided that we obtain initial robust estimates $\hat{\beta}_n^0$ and $\hat{\sigma}_n^0$ of the regression parameters β and standard deviation σ of regression residuals (e.g., by LTS with $h_n = [(n+1)/2] + p$ and by the median absolute deviation), the choice of trimming is done by comparing the empirical distribution function G_n^0 of standardized absolute residuals $|r_i(\hat{\beta}_n^0)/\hat{\sigma}_n^0|$ and the distribution function $F_{|\cdot|}(z) = \Phi(z) - \Phi(-z)$, which describes behavior of $|\varepsilon_i|$ under the assumption of normally distributed errors, $\varepsilon_i \sim N(0,1)$ (G_n^0 is compared to $F_{|\cdot|}$ because LS perform optimally under normality and outliers are extremely improbable under $N(0,1)$). Specifically, Gervini and Yohai (2002) proposed to measure the largest difference between G_n^0 and $F_{|\cdot|}$ in the tail of the distributions,

$$d_n = \sup_{t \geq c} \max\{0, F_{|\cdot|}(t) - G_n^0(t\hat{\sigma}_n^0)\}, \tag{9}$$

where the cut-off point c equals 2.5 and $\hat{\sigma}_n^0 = 1.4826 \cdot \mathrm{MAD}_{i=1,\ldots,n}\, r_i(\hat{\beta}_n^0)$ is the median absolute deviation (MAD) estimate of the residual variance. Using this measure, Čížek (2007a) proposed to estimate using LTS with the following data-dependent choice of trimming: $h_n^a = n - [d_n n]$. The method performs very well under normality, heavy-tailed distributions, and also under heteroscedasticity despite "assuming" the same distribution for all data in (9), see Čížek (2007a).

To employ this idea in the truncated-regression case, we have to take into account that we do not fully observe the regression residuals $y_i^* - x_i^\top \beta$. We can however compute the distribution function $F_{r|\xi}$ of the absolute value of the observed standardized residuals

$|r_{is}(\beta^0)| = |y_i - x_i^\top \beta^0|/\sigma = |\varepsilon_i|/\sigma$ conditional on $\xi_i = x_i^\top \beta^0/\sigma$ if the error term ε_i in the latent model (1) is normally distributed, $N(0,\sigma)$:

$$F_{r|\xi}(t|\xi_i = x_i^\top \beta^0/\sigma = \xi) = \begin{cases} [\Phi(t) - \Phi(-t)]/[1-\Phi(-\xi)] & \text{if } 0 \le t \le \xi, \\ [\Phi(t) - \Phi(-\xi)]/[1-\Phi(-\xi)] & \text{if } |\xi| < t, \\ 0 & \text{if } t < -\xi; \end{cases}$$

(the result follows from the fact that observed error ε_i is truncated at $-\xi_i = -x_i^\top \beta^0$).

Next, we have to average the conditional distribution $F_{r|\xi}$ across ξ, that is, across observations with different values x_i, to obtain an unconditional distribution. To obtain conservative estimates, we will use only observations likely under the assumption $\varepsilon \sim N(0,\sigma)$. In particular, an observation with $\xi_i = x_i^\top \beta^0/\sigma < 0$ in a truncated sample of size n is observed with probability $1 - \Phi^n(|\xi_i|)$ because $\varepsilon_i/\sigma > |\xi_i|$ (Φ^n is the distribution function of $\max_{i=1,\dots,n}[\varepsilon_i/\sigma]$). Considering only observations that can appear in the sample with probability higher than some small $\alpha > 0$, we have to impose condition $1 - \Phi^n(|\xi_i|) > \alpha$, that is, $\xi_i > -\Phi^{-1}(\sqrt[n]{1-\alpha}) = C_n(\alpha)$. The number of observations satisfying this condition will be denoted $n_\alpha = \sum_{i=1}^n I\{\xi_i > C_n(\alpha)\}$. Note that $C_n(\alpha) \to -\infty$ asymptotically and the condition $\xi_i > C_n(\alpha)$ is thus always satisfied for $n \to \infty$.

Finally, denoting the distribution function of $\xi_i = x_i^\top \beta^0/\sigma$ by F_ξ (absolutely continuous by assumption, see Section 4), the cumulative distribution function $F_{r,\alpha}$ of $|r_i(\beta^0)|$ conditional upon $\xi_i = x_i^\top \beta^0/\sigma > C_n(\alpha)$ can be expressed as $F_{r,\alpha}(t) = \int_{C_n(\alpha)}^{+\infty} F_{r|\xi}(t|\xi) dF_\xi(\xi)$. Hence for any $t \ge 0$, it holds that

$$F_{r,\alpha}(t) \quad = \quad \int_{\max\{-t, C_n(\alpha)\}}^{t} \frac{\Phi(t) - \Phi(-\xi)}{1 - \Phi(-\xi)} dF_\xi(\xi) + \int_t^\infty \frac{\Phi(t) - \Phi(-t)}{1-\Phi(-\xi)} dF_\xi(\xi).$$

Given initial estimates $\hat\beta_n^0$ and $\hat\sigma_n^0$ of β^0 and σ and a sample $(x_i, y_i)_{i=1}^n$, an integral

$$\int_A g(t,\xi) dF_\xi(\xi) = \int_{\mathbb{R}} g(t,\xi) I(\xi \in A) dF_\xi(\xi) = \mathsf{E}\{g(t,\xi_i) I(\xi_i \in A)\}$$

is an expectation and can be estimated by the sample mean. Consequently, we can estimate

$F_{r,\alpha}(t)$ by

$$
\begin{aligned}
F_{rn,\alpha}(t) &= \frac{1}{\hat{n}_\alpha}\sum_{i=1}^{n}\frac{\Phi(t)-\Phi(-x_i^\top\hat{\beta}_n^0/\hat{\sigma}_n^0)}{1-\Phi(-x_i^\top\hat{\beta}_n^0/\hat{\sigma}_n^0)} I(\max\{-t,C_n(\alpha)\}<x_i^\top\hat{\beta}_n^0/\hat{\sigma}_n^0<t) \\
&+ \frac{\Phi(t)-\Phi(-t)}{\hat{n}_\alpha}\sum_{i=1}^{n}\frac{1}{1-\Phi(-x_i^\top\hat{\beta}_n^0/\hat{\sigma}_n^0)} I(t\le x_i^\top\hat{\beta}_n^0/\hat{\sigma}_n^0),
\end{aligned} \tag{10}
$$

where $\hat{n}_\alpha = \sum_{i=1}^n I\{x_i^\top\hat{\beta}_n^0/\hat{\sigma}_n^0 > C_n(\alpha)\}$. Analogously to (9), we can then determine the proportion of observations to be trimmed by $h_n^a = n - [d_n n]$ and

$$
d_n = \sup_{t\ge c}\max\{0, F_{rn,\alpha}(t) - G_n^0(t\hat{\sigma}_n^0)\}, \tag{11}
$$

where $c = 2.5$ and $\alpha = 0.001$, for instance.

The only remaining issue is a robust estimate of $\hat{\sigma}_n^0$, which was estimated using MAD in the linear-regression case when responses are fully observed. Since the STLS estimator relies on the symmetry of the latent-error distribution and the usual identification assumption $\mathsf{E}(\varepsilon_i|x_i) = \mathrm{med}(\varepsilon_i|x_i) = 0$, the variance in the truncated case can be consistently estimated using only observations "above" the regression line:

$$
\mathrm{MAD}(\varepsilon_i|x_i) = \mathrm{med}(|\varepsilon_i - \mathrm{med}(\varepsilon_i|x_i)||x_i) = \mathrm{med}(|\varepsilon_i||x_i) = \mathrm{med}(\varepsilon_i|x_i,\varepsilon_i>0),
$$

where the last equality follows from the symmetry of the distribution of ε_i. Using observations with fully observable positive residuals, that is those with $y_i \ge x_i^\top\beta^0 \ge 0$, results in the following estimate of the residual variance:

$$
\hat{\sigma}_n^0 = 1.4826\cdot\mathrm{med}\{r_i(\hat{\beta}_n^0) : y_i \ge x_i^\top\hat{\beta}_n^0 \ge 0\}. \tag{12}
$$

The practical application of the proposed adaptive choice of trimming consists of three steps: (i) an initial estimate $\hat{\beta}_n^0$ and the corresponding residuals $r_i(\hat{\beta}_n^0)$ are obtained by GTE-STLS using trimming constant h_n^0, for example $h_n^0 = [(n+1)/2]+p$; (ii) the residual variance is estimated by (12) and the trimming proportion $h_n^a = n - [d_n n]$ is determined using (11); (iii) the final estimate $\hat{\beta}_n^{(AGTE-STLS)} = \hat{\beta}_n^{(GTE-STLS,h_n^a)}$ is computed using GTE-STLS with the data-dependent trimming constant h_n^a. We refer to this estimate as the adaptive GTE-STLS (AGTE-STLS). The whole procedure is constructed so that AGTE-STLS is asymptotically equivalent to STLS if errors are homoscedastic and normally distributed because the reference distribution $F_r(t) = \int F_{r|\xi}(t|\xi)dF_\xi(\xi)$ in (11) is constructed under the assumption $\varepsilon_i \sim N(0,\sigma)$: hence, both $F_{rn,\alpha}(t) \to F_r(t)$ and $G_n^0(t\hat{\sigma}_n^0) \to F_r(t)$, $t \geq 0$, as $n \to \infty$ if the initial estimates are consistent ($\hat{\beta}_n^0 \to \beta^0$, $\hat{\sigma}_n^0 \to \sigma$, and $C_n(\alpha) = -\Phi^{-1}(\sqrt[n]{1-\alpha}) \to -\infty$ for any fixed $\alpha \in (0,1)$ as $n \to \infty$). This means that AGTE-STLS with adaptive trimming uses all observations "compatible" with normality. Even though outliers are very unlikely under normality, we should make sure that such a data-dependent choice of trimming does not use more observations and improve the variance of estimates at cost of a lower robustness of the whole procedure. The following theorem therefore shows that the breakdown point of AGTE-STLS is not smaller than the breakdown of the initial robust estimator.

Theorem 2 *Let $(y_i, x_i)_{i=1}^n$ be a sequence of independent and identically distributed random vectors, which are almost surely in a general position for $n > 2p$. Further, let ϵ_n^{0*} denote the finite-sample breakdown point of an initial estimator $(\hat{\beta}_n^0, \hat{\sigma}_n^0)$ of regression parameters and residual variance. Then the finite-sample breakdown point of the AGTE-STLS estimator is larger than or equal to ϵ_n^{0*} if the GTE-STLS estimators with trimming constants h_n are identified for all $[(n+1)/2] + p \leq h_n \leq n$.*

(We will see in Section 4 that the identification assumptions of GTE-STLS are identical to those for STLS discussed in Powell (1986).)

3.3 Robust estimation of censored regression

In Sections 3.1 and 3.2, we introduced a robust (A)GTE-STLS estimator of truncated regression. Using the same strategy, combining directly GTE and SCLS, does not however seem to be feasible in the censored regression model. Intuitively, the GTE type of trimming would invalidate the conditional mean or median restrictions. For example, if ε^* conditional on x is symmetrically distributed around 0 on the whole real line and $\varepsilon = \max\{\varepsilon^*, -x^\top\beta\}$, SCLS relies on $\mathsf{E}[\min\{\varepsilon, x^\top\beta\}I(0 < x^\top\beta)|x] = 0$ (Powell, 1986), which does not hold once large values of $|\varepsilon|$ are trimmed: $\mathsf{E}[\min\{\varepsilon, x^\top\beta\}I(|\varepsilon| \leq K)I(0 < x^\top\beta)|x] \neq 0$ for any $K > x^\top\beta$. On the other hand, the proposed (A)GTE-STLS estimator can be directly applied in censored regression if only non-censored observations are used. Specifically, having a censored sample $(x_i, y_i)_{i=1}^n$, we can estimate the parameters of the model CM by applying AGTE-STLS to the subsample $(x_i, y_i)_{y_i>0}$ of the original data. Such estimates will have a positive breakdown point proportional to $[(n - c_n)/(2n)]$, where $c_n = \sum_{i=1}^{n} I(y_i = 0)$ denotes the number of censored observations, because GTE-STLS can trim at most half of the $n - c_n$ observations contained in the truncated sample $(x_i, y_i)_{y_i>0}$.

Despite the adaptive choice of trimming, the application of AGTE-STLS in censored regression is suboptimal since it neglects the information present in the censored observations with $y_i = 0$. Since the (A)GTE-STLS provides already a robust estimator of censored regression, we can however employ the so-called one-step estimation, which is often used for robust M-estimators of linear regression (Simpson et al., 1992; Welsh and Ronchetti, 2002). Since nonlinear methods such as M-estimators are computed using iterative optimization techniques, the one-step estimation employs an initial robust estimate $\hat{\beta}_n^0$ as a starting point to perform a single step of the iterative optimization procedure. The resulting estimator can then inherit robust properties of the initial estimator and asymptotic properties of the M-estimator (Simpson et al., 1992).

In the case of SCLS, the iterative computation algorithms were proposed by Powell (1986) and Santos Silva (2001). We discuss here only the first one, since the latter Newton-type algorithm is less stable and cannot preserve the robust properties of AGTE-STLS at an arbitrary sample. Powell's algorithm relies on the moment condition $\mathsf{E}[x_i \min\{\varepsilon_i, x_i^\top \beta^0\} I(x_i^\top \beta^0 > 0)] = 0$, which holds under the conditional symmetry of ε_i given x_i. Substituting $\varepsilon_i = y_i - x_i^\top \beta$, we can solve this moment condition for β and replacing β^0 by an initial estimate $\hat{\beta}_n^0$ then leads to the following estimate $\hat{\beta}_n^C$ (cf. Powell, 1986, equation (2.13))

$$\hat{\beta}_n^C(\hat{\beta}_n^0) = \left[\frac{1}{n}\sum_{i=1}^{n} x_i x_i^\top I(x_i^\top \hat{\beta}_n^0 > 0)\right]^{-1} \frac{1}{n}\sum_{i=1}^{n} x_i \min\{y_i, 2x_i^\top \hat{\beta}_n^0\} I(x_i^\top \hat{\beta}_n^0 > 0). \tag{13}$$

Note that the inverted matrix has to be non-singular if SCLS is identified (see Section 4 and Powell, 1986, Assumption R). If the inverted matrix is singular, we just define $\hat{\beta}_n^C(\hat{\beta}_n^0) = \hat{\beta}_n^0$. Using the iterative formula (13) and (A)GTE-STLS as the initial estimator $\hat{\beta}_n^0$, we can now propose robust one-step SCLS estimators of the censored regression model CM:

$$\hat{\beta}_n^{(ONE-SCLS,h_n)} = \hat{\beta}_n^C(\hat{\beta}_n^{(GTE-STLS,h_n)}), \tag{14}$$

$$\hat{\beta}_n^{(ONE-SCLS,A)} = \hat{\beta}_n^C(\hat{\beta}_n^{(AGTE-STLS)}); \tag{15}$$

the resulting estimators are referred to as ONE-SCLS.

Even though the one-step estimation seems to be necessary only in the context of censored regression, we can also use it in the context of truncated regression as an alternative to AGTE-STLS. The STLS estimator relies on the moment condition $\mathsf{E}[\varepsilon_i x_i I(\varepsilon_i < x_i^\top \beta^0)] = 0$, which holds under the conditional symmetry of ε_i given x_i. Substituting $\varepsilon_i = y_i - x_i^\top \beta$, solving this moment condition for β, and replacing β^0 by an initial estimate $\hat{\beta}_n^0$ then produces the following iterative equation

$$\hat{\beta}_n^T(\hat{\beta}_n^0) = \left[\frac{1}{n}\sum_{i=1}^n x_i x_i^\top I(y_i < 2x_i^\top \hat{\beta}_n^0)\right]^{-1} \frac{1}{n}\sum_{i=1}^n x_i y_i I(y_i < 2x_i^\top \hat{\beta}_n^0). \tag{16}$$

We again have to assume that the STLS estimator is identified and that the inverted matrix is thus non-singular (see Section 4 and Powell, 1986, Assumption R and E2); otherwise, $\hat{\beta}_n^T(\hat{\beta}_n^0) = \hat{\beta}_n^0$. Using formula (16) and (A)GTE-STLS as the initial estimator $\hat{\beta}_n^0$, we can propose robust one-step STLS estimators of the truncated regression model TM defined by

$$\hat{\beta}_n^{(ONE-STLS,h_n)} = \hat{\beta}_n^T(\hat{\beta}_n^{(GTE-STLS,h_n)}), \tag{17}$$

$$\hat{\beta}_n^{(ONE-STLS,A)} = \hat{\beta}_n^T(\hat{\beta}_n^{(AGTE-STLS)}); \tag{18}$$

the method is referred to as ONE-STLS.

For both truncated and censored regression, we thus propose to use the robust GTE-STLS or AGTE-STLS estimators as the initial estimators and to perform one step of the iterative STLS or SCLS algorithm, (16) or (13). On one hand, the resulting ONE-STLS and ONE-SCLS estimators will always use all observations and thus more information from data. On the other hand, the one-step estimation preserves the breakdown point of the initial estimator as we show in the following theorem.

Theorem 3 *Let $(y_i, x_i)_{i=1}^n$ be a sequence of independent and identically distributed random vectors, which are almost surely in a general position for $n > 2p$. Further, let ϵ_n^{0*} denote the finite-sample breakdown point of an initial estimator $\hat{\beta}_n^0$ of regression parameters. Then the finite-sample breakdown points of the ONE-STLS and ONE-SCLS estimators are larger than or equal to ϵ_n^{0*}.*

Despite the same breakdown point of AGTE-STLS, ONE-STLS, and ONE-SCLS, it seems that AGTE-STLS will be less sensitive, in terms of bias, to outliers and data contamination because it can reject some observations from its objective function. On the other hand, ONE-STLS and ONE-SCLS could possibly exhibit smaller variances of estimates because they always use all available observations. A detailed comparison of all estimators by means of asymptotic properties and by means of Monte Carlo simulations follows in Sections 4 and 5.

4 Asymptotic properties

Let us now analyze the asymptotic properties of the proposed (A)GTE-STLS, ONE-STLS, and ONE-SCLS estimators. After introducing notation and assumptions needed for the asymptotic analysis, we will first find the asymptotic distribution of GTE-SCLS with the trimming constant being a fixed fraction of the sample, $h_n = [\lambda n]$, where $\lambda \in (1/2, 1\rangle$. Later, we look at the asymptotic behavior of AGTE-SCLS, that is, GTE-SCLS with the adaptively chosen trimming, and finally, we study the asymptotic distributions of the proposed one-step estimators.

First, the (unconditional) distribution function of ε_i in model (1) is referred to as F and its density function is denoted f, provided that it exists. A density function f of a random variable with zero mean will be called (strictly) unimodal if $f(z_1) \geq f(z_2)$ ($f(z_1) > f(z_2)$) for any $|z_1| \geq |z_2|$. Further, let us introduce the concept of β-mixing, which is central to the distributional assumptions made here. A sequence of random variables $\{x_i\}_{i\in\mathbb{N}}$ is said to be absolutely regular (or β-mixing) if $\beta_m = \sup_{i\in\mathbb{N}} \mathsf{E}\{\sup_{B\in\sigma^f_{i+m}} |P(B|\sigma^p_i) - P(B)| \to 0\}$ as $m \to \infty$, where σ-algebras $\sigma^p_i = \sigma(x_i, x_{i-1}, \ldots)$ and $\sigma^f_i = \sigma(x_i, x_{i+1}, \ldots)$; see Davidson (1994) for details. Numbers β_m, $m \in \mathbb{N}$, are called mixing coefficients.

Now, let us present the assumptions used to derive the asymptotic distribution of all proposed estimators.

Assumption A

A1 Random vectors $(x_i, y_i)_{i\in\mathbb{N}}$ form a weakly stationary absolutely regular sequence with mixing coefficients β_m satisfying $m^{r/(r-2)}(\log m)^{2(r-1)/(r-2)}\beta_m \to 0$ as $m \to \infty$ for some $r > 2$ and have finite rth moments. Moreover, let $\mathsf{E}\{x_i x_i^\top I(x_i^\top \beta^0 \geq \gamma_0)\} = Q$ be a positive definite matrix and $\mathsf{E}\,\|x_i\|^{4+\delta} < \infty$ for some $\gamma_0 > 0$ and $\delta > 0$.

A2 Let $\{\varepsilon_i\}_{i\in\mathbb{N}}$ be a sequence of independently distributed random variables, and conditional on x_i, let $\{\varepsilon_i\}_{i\in\mathbb{N}}$ be symmetrically distributed with finite second moments, $\mathsf{E}(\varepsilon_i|x_i) = 0$ and $\mathsf{var}(\varepsilon_i|x_i) < \infty$. The conditional distribution function F_i of ε_i given x_i is assumed to be absolutely continuous with its probability density function f_i being bounded, positive at 0, and continuously differentiable uniformly in $i \in \mathbb{N}$. Additionally, the error densities f_i have to be unimodal and strictly unimodal in some neighborhood of zero uniformly in $i \in \mathbb{N}$.

A3 The true parameter β^0 lies in the interior of a compact parametric space B.

The majority of Assumptions A correspond to Assumptions P, R, E1, and E2 used by Powell (1986) to analyze the asymptotic properties of STLS and SCLS. Compared to Powell (1986), we allow for dependence in data (Assumption A1), but additionally require the differentiability of the error density functions. Let us also note that, while the moment assumptions might look relatively strong on the first look, they can be weakened in some cases: (i) if only consistency of (A)GTE-STLS or ONE-SCLS is required, finite $(2+\delta)$th moments are sufficient and (ii) if (A)GTE-STLS is considered and there is a positive amount of trimming, $h_n = [\lambda n] < n$ and $\lambda < 1$, only trimmed moments of all variables have to exist (cf. Čížek, 2007b). Finally, note that the assumption of random carriers for all variables is made for the sake of simplicity and the results apply in the presence of deterministic variables as well.

Let us first derive the asymptotic distribution of (A)GTE-STLS.

Theorem 4 *Let Assumption A hold. Then the GTE-STLS estimator $\hat{\beta}_n^{(GTE-STLS,h_n)}$ of the model TM using trimming $h_n = [\lambda n]$, $\lambda \in (1/2, 1\rangle$, is consistent and asymptotically normal, that is,*

$$\sqrt{n}\left(\hat{\beta}_n^{(GTE-STLS,h_n)} - \beta^0\right) \xrightarrow{F} N(0, V(\lambda))$$

as $n \to +\infty$. Furthermore, the asymptotic distribution of GTE-STLS does not change if h_n^a is random (data-dependent) and $h_n^a/n \to \lambda$ in probability as $n \to +\infty$.

An important consequence of Theorem 4 is that the asymptotic distribution does not depend on (random) trimming h_n^a as long as $h_n^a/n \to \lambda$ in probability. Therefore, the asymptotic distribution of AGTE-STLS proposed in Section 3.2 is the same as for GTE-STLS with trimming $[n \lim_{n\to\infty}(h_n^a/n)]$. In particular, if the latent error term in (1) is homoscedastic and normally distributed, $h_n^a/n = (n - [d_n n])/n \to 1$ as $n \to \infty$ as explained in Section 3.2 and the asymptotic distribution of AGTE-STLS is identical to that of STLS.

Next, although we proved the asymptotic normality of (A)GTE-STLS, we do not specify the precise form of the asymptotic variance matrix $V(\lambda)$. Even though it could be formally derived, it does not have a computationally feasible form, especially under heteroscedasticity (cf. Powell, 1986, and the nontrivial asymptotic distribution of STLS itself). Hence, it has to be computed by a parametric or a robust nonparametric bootstrap, for instance (e.g., Hall and Presnell, 1999; Salibian-Barrera and Zamar, 2002).

Alternatively, one can use the GTE-STLS estimate only as a starting point and compute the one-step ONE-STLS or ONE-SCLS estimators. Although this could possibly increase bias due to outliers and data contamination, ONE-STLS and ONE-SCLS employ all information in a sample, and more importantly, are asymptotically first-order equivalent to STLS and SCLS as we show in the following theorem.

Theorem 5 *Let Assumption A hold and let $\hat{\beta}_n^0$ be a $\sqrt{n}$-consistent estimate of parameters β in the models TM and CM. Then the ONE-STLS estimator $\hat{\beta}_n^{(ONE-STLS)} = \hat{\beta}_n^T(\hat{\beta}_n^0)$ of the model TM based on $\hat{\beta}_n^0$ and the ONE-SCLS estimator $\hat{\beta}_n^{(ONE-SCLS)} = \hat{\beta}_n^C(\hat{\beta}_n^0)$ of the model CM based on $\hat{\beta}_n^0$ are first-order asymptotically equivalent to the STLS estimate $\hat{\beta}_n^{(STLS)}$ and to the SCLS estimate $\hat{\beta}_n^{(SCLS)}$, respectively:*

$$\sqrt{n}(\hat{\beta}_n^{(ONE-STLS)} - \hat{\beta}_n^{(STLS)}) \to 0 \qquad \text{and} \qquad \sqrt{n}(\hat{\beta}_n^{(ONE-SCLS)} - \hat{\beta}_n^{(SCLS)}) \to 0$$

in probability as $n \to \infty$.

Theorem 5 documents the well-known fact that the asymptotic distribution of one-step estimators can be independent of the initial estimator. It also means that the proposed one-step estimators, ONE-STLS and ONE-SCLS, have the same asymptotic distributions as STLS and SCLS and the asymptotic variances by Powell (1986) can thus be applied as long as the number of observations is sufficiently large. See Section 5 for more details.

5 Finite-sample properties

In this section, we present a Monte Carlo study done to assess finite-sample behavior of the proposed GTE-STLS with $h_n = [(n+1)/2] + p$, AGTE-STLS, ONE-STLS, and ONE-SCLS estimators; note that ONE-STLS-0 and ONE-SCLS-0 will denote the one-step estimators based on GTE-STLS, whereas ONE-STLS-A and ONE-SCLS-A will refer to estimates using AGTE-STLS as the initial estimator. The proposed estimators are compared with MLE, STLS, and SCLS in the context of truncated regression (Section 5.1) and censored regression (Section 5.2).

In all cases, we generate data using a latent model (1), $y_i^* = x_i^\top \beta + \varepsilon_i$, and subsequently, we omit the observations with $y_i^* < 0$ in the case of the truncated model TM or we set $y_i = \max\{y_i^*, 0\}$ in the case of the censored model CM. The presented simulations, based on 1000 simulated samples, are done for sample sizes $n = 100$, 200, and 400 using $\beta = (1, -1, 1)^\top$ and ε_i from various distributions (results however do not qualitatively change with the dimension of β). In this setup, for example, $n = 100$ and $x_{1i}, x_{2i}, \varepsilon_i \sim N(0, 1)$ lead to samples than contain 17–40 observations with $y_i^* \leq 0$ with probability more than 99%. We use the following data generating processes, where $x_{1i} \sim N(0, 1)$ and $x_{2i} \sim N(0, 1)$ unless stated otherwise:

NORM: Clean Gaussian data, $\varepsilon_i \sim N(0, 1)$.

DEXP: Data with errors following a double-exponential distribution, $\varepsilon_i \sim DExp(1)$.

STD(d): Data with errors from a heavy-tailed distribution, $\varepsilon_i \sim t(d)$, where $t(d)$ denotes the Student distribution with d degrees of freedom.

HETX: Data with heteroscedastic errors, $\varepsilon_i \sim N(0, e^{2x_1})$, where variance depends on covariate values.

HETZ: Data with heteroscedastic errors, $\varepsilon_i \sim N(0, z)$, where variance depends on unobservable $z \sim U(0.25, 4)$ and $U(a, b)$ denotes the uniform distribution on (a, b).

OUT($a; l_1, l_2$): Data contaminated by $[an], a \geq 0$, outliers at location $(l_1, l_2) \in \mathbb{R}^2$. Specifically, a fraction $1 - a$ of observations satisfies $x_{1i} \sim N(0, 1)$, $x_{2i} \sim N(0, 1)$, and $\varepsilon_i \sim N(0, 1)$, whereas the complementary fraction a of remaining observations follows $x_{1i} \sim N(l_1, 1)$, $x_{2i} \sim N(l_2, 1)$, and $\varepsilon_i \sim U(-50, 50)$.

Table 1: The absolute bias and MSE of various estimators for truncated data NORM with $n = 100$, 200, and 400 observations.

Estimation method	$n = 100$		$n = 200$		$n = 400$	
	Bias	MSE	Bias	MSE	Bias	MSE
MLE	0.020	0.053	0.009	0.024	0.005	0.012
STLS	0.028	0.086	0.019	0.044	0.006	0.020
GTE-STLS	0.073	0.283	0.030	0.167	0.042	0.096
AGTE-STLS	0.042	0.108	0.017	0.050	0.011	0.021
ONE-STLS-0	0.018	0.113	0.016	0.061	0.016	0.034
ONE-STLS-A	0.021	0.090	0.012	0.044	0.007	0.020

(Sample size spans all columns.)

Let us note here that we generally look at four types of data containing outlying observations (assuming that $l > 1$ now): (i) $\mathrm{OUT}(a; 0, 0)$, where outliers are located at the same place as the remaining observations; (ii) $\mathrm{OUT}(a; l, l)$, where outliers are on average at distance $\sqrt{2}l$ from the remaining observations, but they are all close to the censoring hyperplane $((1, l, l)^\top(1, -1, 1) = 1)$; (iii) $\mathrm{OUT}(a; -l, l)$, where outliers are on average at distance $\sqrt{2}l$ from the remaining observations and they are in the region without (or with a very small number of) censored or truncated observations, at least for larger values of l $((1, -l, l)^\top(1, -1, 1) = 1 + 2l > 0)$; and (iv) $\mathrm{OUT}(a; l, -l)$, where outliers are on average at distance $\sqrt{2}l$ from the remaining observations and they are in the region with many censored or truncated observations, at least for larger values of l $((1, l, -l)^\top(1, -1, 1) = 1 - 2l < 0)$.

To judge the finite-sample behavior of the discussed estimators, we compare different estimators by means of the absolute bias $\| \mathsf{E}\, \hat{\beta}_n^{(T)} - \beta\|$ and by means of the squared error $\|\hat{\beta}_n^{(T)} - \beta\|^2$, where $\hat{\beta}_n^{(T)}$ represents the estimate by a method T. Because we also analyze the behavior of all methods in the presence of outliers and the number of simulations is relatively limited ($S = 1000$), we estimate $\mathsf{E}\, \hat{\beta}_n^{(T)}$ in the absolute bias by $\mathrm{med}_{s=1,\ldots,S}\, \hat{\beta}_n^{(T,s)}$ and further report the median squared error (MSE) $\mathrm{med}_{s=1,\ldots,S} \|\hat{\beta}_n^{(T,s)} - \beta\|^2$ rather than more usual mean squared error ($\hat{\beta}_n^{(T,s)}$ denotes the T estimate for simulated sample s). In some cases, we also compute the quartiles of the squared error (QSE), which are more informative than just MSE.

5.1 Truncated regression

Let us first discuss the simulation results for the truncated model TM, which are summarized in Tables 1 (data NORM), 2 (data NORM, DEXP, STD(5), HETX, and HETZ), and 3 (data $\mathrm{OUTLIER}(a; l_1, l_2)$).

Table 2: The MSE of various estimators for truncated samples without outliers using $n = 200$ observations.

Estimation method	Data generating process				
	NORM	DEXP	STD(5)	HETX	HETZ
MLE	0.024	0.093	0.060	2.947	1.756
STLS	0.044	0.046	0.053	0.030	0.152
GTE-STLS	0.167	0.050	0.147	0.032	0.121
AGTE-STLS	0.050	0.043	0.055	0.022	0.139
ONE-STLS-0	0.061	0.035	0.063	0.022	0.095
ONE-STLS-A	0.044	0.042	0.052	0.021	0.136

Table 3: The MSE of various estimators for truncated samples with 10% outliers using $n = 200$ observations.

Estimation method	Data generating process			
	$\mathrm{OUT}(0.1; 0, 0)$	$\mathrm{OUT}(0.1; 8, 8)$	$\mathrm{OUT}(0.1; -8, 8)$	$\mathrm{OUT}(0.1; 8, -8)$
MLE	($> 10^9$, $> 10^{10}$)	(14.17, 20.39)	(6.385, 10.86)	(30.94, 40.32)
STLS	(0.034, $> 10^4$)	($> 10^2$,$> 10^5$)	($> 10^3$, $> 10^5$)	($> 10^2$, $> 10^5$)
GTE-STLS	(0.071, 0.319)	(0.077, 0.310)	(0.072, 0.331)	(0.068, 0.303)
AGTE-STLS	(0.021, 0.110)	(0.028, 0.224)	(0.052, 0.730)	(0.022, 0.123)
ONE-STLS-0	(0.026, 0.134)	(0.028, 0.148)	(0.080, 0.524)	(0.029, 0.133)
ONE-STLS-A	(0.020, 0.104)	(0.027, 0.223)	(0.110, 1.130)	(0.022, 0.123)

In Table 1, we summarize the absolute bias and MSE of all truncated-regression estimators for data NORM and sample sizes $n = 100$, 200, and 400. As all methods provide consistent estimates and the biases are thus approximately zero, we discuss primarily MSE. Obviously, MLE provides the efficient estimates in this case. The second best estimator is STLS because all other (robust) methods directly or indirectly employ trimming of observations. On one hand, the initial robust estimator, GTE-STLS, exhibits largest bias and MSE of all methods. On the other hand, both AGTE-STLS and ONE-STLS perform significantly better than the initial robust estimator. The ONE-STLS-A actually matches the performance of STLS at all sample sizes and AGTE-STLS does so for $n = 400$. The relatively worse performance of AGTE-STLS for small samples is likely related to the precision of the residual variance estimate $\hat{\sigma}_n^0$, which uses only half of the sample – see (12). Keeping in mind that the relative performance of AGTE-STLS is worse for $n = 100$ and better for $n = 400$, we present the following results only for $n = 200$.

Next, we compare performance of all methods under various distributional models, see Table 2. Clearly, MLE is no longer the best estimator: it exhibits the largest MSE for data DEXP and it is inconsistent under heteroscedasticity. Similarly, STLS is preferable only for data NORM and STD(5), but is inferior to the robust estimators in the presence of heteroscedasticity or exponentially distributed errors. Interestingly, GTE-STLS has a large MSE for data NORM and STD(5), but performs well for data DEXP, HETX, and HETZ (in the latest case, it actually outperforms AGTE-STLS). In the cases when GTE-STLS performs well, the ONE-STLS-0 estimator is the best one. In the other two cases, AGTE-STLS and ONE-STLS-A, which have similarly large MSEs, are better than ONE-STLS-0.

Finally, the behavior of all estimator is analyzed for data containing 10% of outliers, data OUT($a; l_1, l_2$) for $a = 0.1$, $l_j \in \{-8, 0, 8\}$ and $j = 1, 2$ (the results for the contamination levels $a = 0.05$ and $a = 0.20$ are very similar). Because the influence of contaminated observations can substantially vary with their precise location and the magnitude of the outlying observations, we report in this case the first and third quartiles of the squared estimation errors (QSEs) instead of MSE, see Table 3. Clearly, both MLE and STLS are in this case extremely influenced by outlying observations irrespective of their location. On the other hand, the most robust GTE-STLS estimator exhibits relatively small QSE, which are stable irrespective of the type of contamination. Other robust alternatives, AGTE-STLS and ONE-STLS, improve upon the initial GTE-STLS estimator except for data OUT(0.1; -8, 8), which contain outlying leverage points in the area with a low truncation probability. The overall best performance can be attributed here to ONE-STLS-0, which is the best of the adaptive and one-step methods for data OUT(0.1; -8, 8) and closely matches the alternatives under other types of contamination.

Altogether, while STLS is sensitive to data contamination, GTE-STLS is robust estimator of truncated regression with very stable, but less precise results across various data-generating models. Considering the adaptive and one-step alternatives, ONE-STLS-0 seems to be the most universal estimator since it performs very well in most of non-Gaussian situations and it stays further behind AGTE-STLS and ONE-STLS-A only for clean Gaussian data.

5.2 Censored regression

Now, we will look at the simulation results for the censored model CM, which are summarized in Tables 4 (data NORM), 5 (data NORM, DEXP, STD(5), HETX, and HETZ), and 6 (data OUTLIER($a; l_1, l_2$)).

In Table 4, the results concerning all discussed censored-regression estimators are sum-

Table 4: The absolute bias and MSE of various estimators for censored data NORM with $n = 100$, 200, and 400 observations.

Estimation method	Sample size $n = 100$ Bias	MSE	$n = 200$ Bias	MSE	$n = 400$ Bias	MSE
MLE	0.005	0.033	0.005	0.015	0.004	0.007
SCLS	0.015	0.055	0.010	0.025	0.003	0.013
GTE-STLS	0.107	0.297	0.062	0.170	0.020	0.097
AGTE-STLS	0.062	0.145	0.015	0.067	0.004	0.030
ONE-SCLS-0	0.041	0.091	0.038	0.049	0.013	0.025
ONE-SCLS-A	0.047	0.079	0.023	0.036	0.009	0.018

Table 5: The MSE of various estimators for censored samples without outliers using $n = 200$ observations.

Estimation method	Data generating process NORM	DEXP	STD(5)	HETX	HETZ
MLE	0.015	0.028	0.024	0.307	0.080
SCLS	0.025	0.038	0.036	0.023	0.111
GTE-STLS	0.169	0.067	0.148	0.033	0.210
AGTE-STLS	0.067	0.059	0.083	0.031	0.237
ONE-SCLS-0	0.049	0.038	0.055	0.023	0.119
ONE-SCLS-A	0.036	0.040	0.050	0.023	0.153

marized for data NORM and sample sizes $n = 100$, 200, and 400. As all methods provide consistent estimates, the biases are approximately zero. Comparing MSEs, MLE provides the efficient estimates in this case. The second best estimator is SCLS because, similarly to truncated regression, all other (robust) methods directly or indirectly employ trimming of observations. The largest bias and MSE can be attributed to the initial robust estimator GTE-STLS – this comes as no surprise as it uses only half of non-censored observations, that is, about 30–40% observations. On the other hand, both AGTE-STLS and ONE-SCLS perform significantly better than GTE-STLS, with ONE-SCLS-A being best followed by ONE-SCLS-0 and by AGTE-STLS, which uses only non-censored observations and which is thus worst. Note that the MSE of ONE-SCLS-0 (ONE-SCLS-A) is roughly twice (by 38–45%) larger than that of SCLS, but this difference should asymptotically converge to zero (Theorem 5). This indicates that using the asymptotic covariance matrix of SCLS with ONE-SCLS in small samples can lead to misleading inference and that bootstrap might be preferable in this situation. Finally, given that these qualitative results do no change much with the sample size, the following results are presented only for $n = 200$.

Table 6: The MSE of various estimators for censored samples with 10% outliers using $n = 200$ observations.

Estimation method	Data generating process			
	OUT$(0.1; 0, 0)$	OUT$(0.1; 8, 8)$	OUT$(0.1; -8, 8)$	OUT$(0.1; 8, -8)$
MLE	(9.621, 26.13)	(17.32, 33.61)	(8.141, 15.83)	(11.45, 19.57)
SCLS	(0.029, 0.321)	(0.160, $> 10^5$)	($> 10^3$, $> 10^6$)	(0.049, $> 10^4$)
GTE-STLS	(0.075, 0.337)	(0.071, 0.346)	(0.081, 0.374)	(0.072, 0.312)
AGTE-STLS	(0.032, 0.155)	(0.037, 0.324)	(0.058, 0.627)	(0.031, 0.217)
ONE-SCLS-0	(0.030, 0.150)	(0.035, 0.168)	(0.193, 0.758)	(0.023, 0.108)
ONE-SCLS-A	(0.025, 0.119)	(0.030, 0.240)	(0.216, 0.996)	(0.021, 0.124)

Next, we compare performance of all methods under various distributional models, see Table 5. MLE is still the best estimator for models without heteroscedasticity. SCLS is slightly worse than MLE, but it is not biased under heteroscedasticity (data HETX). Let us now consider the robust estimators. Similarly to the truncated-regression case, GTE-STLS has larger MSEs for data NORM and STD(5) and smaller MSEs for data DEXP, HETX, and HETZ relative to other estimators. In the cases when GTE-STLS performs well, the ONE-SCLS-0 estimator matches or outperforms ONE-SCLS-A and AGTE-STLS. On the other hand, ONE-SCLS-A provides practically the best robust estimates except for data HETZ and its MSEs are 0–50% larger than those of SCLS; this difference further reduces to -6–37% for $n = 400$.

Finally, the behavior of all estimators is analyzed for data containing 10% of outliers, data OUT$(a; l_1, l_2)$ for $a = 0.1$, $l_j \in \{-8, 0, 8\}$ and $j = 1, 2$ (the results are qualitatively similar to those for $a = 0.05$ and $a = 0.20$). Because the influence of contaminated observations can substantially vary with their precise location, the magnitude of the outlying observations, and the number of censored and non-censored outliers, we again report the first and third QSE instead of MSE, see Table 6. Clearly, MLE is in this case extremely influenced by outlying observations irrespective of their location. Additionally, SCLS can withstand vertical outliers, data OUT(0.1; 0, 0), but fails in all other data containing outliers ($l_1 \neq 0$ and $l_2 \neq 0$). In other words, SCLS is not influenced by contaminated observations only if they are not outlying in the space of explanatory variables. (The low values of the first QSE in some models is a result of the fact that many outliers can be censored in some simulated data; the estimated median biases of SCLS are however always large: 6–80.) On the other hand, the most robust GTE-STLS estimator exhibits relatively small QSE, which are stable irrespective of the type of contamination. Similarly to the truncated-regression case, other robust alternatives, AGTE-STLS and ONE-SCLS, improve upon the initial GTE-STLS estimator except for data OUT(0.1; -8, 8). Moreover contrary to MLE and SCLS, all robust estimators exhibit smaller QSE as the sample size increases; for example, the QSE of all robust estimators are 35–50% smaller for $n = 400$ than for $n = 200$ (Table 6). The overall best performance could be probably attributed to ONE-SCLS-0, although the difference between ONE-SCLS-0 and ONE-SCLS-A is not so pronounced as in the truncated case and ONE-SCLS-A becomes preferable with an increasing sample size.

Altogether, SCLS was shown to be sensitive to data contamination, whereas GTE-STLS provided consistent, though less precise estimates across all data-generating models. Selecting from the adaptive and one-step robust alternatives, ONE-SCLS-A is a preferable robust method in most considered models, especially in models with homoscedastic errors, although it deals quite well also with heteroscedastic and contaminated data in larger samples. However if one believes that data are small, exhibit heteroscedasticity, or contain many outlying observations, ONE-SCLS-0 is a better choice.

6 Conclusion

In this work, we introduced new semiparametric high breakdown-point estimators of truncated and censored regression models. Being derived from STLS and SCLS, the estimators are consistent under weak identification assumptions, are asymptotically normally distributed, and additionally, the one-step estimators are asymptotically equivalent to the original STLS and SCLS. Finite sample performance of the proposed robust estimators matches that of STLS in the case of truncated regression. There is a difference in the variance of the robust and SCLS estimates in the case of censored regression, but it is relatively limited even in small samples and more than enough compensated for by the robust properties of the proposed methods.

The robust estimation of truncated and censored regression is studied here only for STLS and SCLS in simple limited-dependent-variable models without, for example, panel-data structures or endogenous regressors. The extension of the GTE and one-step estimation concepts to other estimation methods and models is relatively straightforward and can mimic the corresponding extensions of STLS and SCLS, for instance, but remains a topic for further research.

Appendix

Proof of Theorem 1: Consider a sample $Z_n = (x_i, y_i)_{i=1}^n$, a constant $K > 0$, and $2p$ contaminated observations $(\tilde{x}_j, \tilde{y}_j)_{j=1}^{2p}$. For $j = 1, \ldots, 2p$, the contaminated observations $(\tilde{x}_j, \tilde{y}_j)$ have values $\tilde{x}_j = (1, (-1)^j s e_{[j/2]}^\top)^\top$ and $\tilde{y}_j = K$, where the first element of $\tilde{x}_j$ represents the intercept, $e_j = (0, \ldots, 0, 1, 0, \ldots, 0)^\top$ denotes the jth basis vector of the Euclidean space $\mathbb{R}^{p-1}$, and $s \in \mathbb{N}$. Let Z_n^s denote the sample created from Z_n by replacing its first $2p$ observation by data points $(\tilde{x}_j, \tilde{y}_j)_{j=1}^{2p}$ and let $\mu(Z) = \sum_{(x_i,y_i) \in Z} y_i / |Z|$ denote the mean of responses in a sample Z. We now prove that the slopes of the MLE and STLS estimates $\hat{\beta}_n^{(MLE-TM)}(Z_n^s)$ and $\hat{\beta}_n^{(STLS)}(Z_n^s)$ in the truncated regression model TM and the MLE and SCLS estimates $\hat{\beta}_n^{(MLE-CM)}(Z_n^s)$ and $\hat{\beta}_n^{(SCLS)}(Z_n^s)$ in the censored regression model CM evaluated at the sequence of samples $Z_n^s, s \in \mathbb{N}$, converge to 0 as $s \to \infty$ and that the estimated intercepts converge to $+\infty$.

First, note that the objective functions of all estimators evaluated at $\beta_C = (C, 0, \ldots, 0)$ are finite for $C \geq 0$ at any $Z_n^s, s \in \mathbb{N}$, see equations (2)–(5). On the other hand, if we consider any β with nonzero slope coefficients, $\left\| \beta - (\beta_1, 0 \ldots, 0)^\top \right\| \neq 0$, and samples Z_n^s, the residuals $\tilde{y}_j - \tilde{x}_j^\top \beta$ are equal to $K - (-1)^j s \beta_{1+[j/2]}$, and consequently, at least one residual converges to $+\infty$ as $s \to \infty$ (the residuals $y_i - x_i^\top \beta$ of the remaining observations that are not contaminated are finite and independent of s). Consequently, the objective function of all estimators converges to $+\infty$ as $s \to \infty$ for any β with nonzero slope coefficients and the all estimates thus must asymptotically have the form $\beta_C = (C, 0, \ldots, 0)$ as $s \to \infty$.

Next, the MLE, STLS, and SCLS estimators at Z_n^s reduce thus to the least squares as $s \to \infty$ if K is sufficiently large because $0 \leq \beta_C^\top x$ and $y_i/2 \leq \beta_C^\top x$ for any $x \in \mathbb{R}^p$ and $\max_{i=1,\ldots,n} y_i/2 \leq K$. These estimates will thus equal $\beta_\mu = (\mu(Z_n^s), 0, \ldots, 0)$ at Z_n^s for $s \to \infty$. Letting $K \to +\infty$ (e.g., by setting $K = \sqrt{s}$) then results in $\mu(Z_n^s) \to +\infty$ and we can conclude that $\|\hat{\beta}_n^{(MLE-TM)}(Z_n^s)\| \to \infty$, $\|\hat{\beta}_n^{(STLS)}(Z_n^s)\| \to \infty$, $\|\hat{\beta}_n^{(MLE-CM)}(Z_n^s)\| \to \infty$, and $\|\hat{\beta}_n^{(SCLS)}(Z_n^s)\| \to \infty$ as $s \to \infty$ and $K \to \infty$.

Hence, we have shown that there is a sequence of samples Z_n^s contaminated by $2p$ observations such that the norms of the MLE, STLS, and SCLS estimators converge to ∞ (and their slopes break down to 0) as $s \to \infty$ irrespective of the initial sample Z_n, their breakdown point is therefore bounded from above by $2p/n$, and asymptotically, $2p/n \to 0$. $\square$

Proof of Theorem 2: For a given sample $\{y_i, x_i\}_{i=1}^n$ of size n, let $\epsilon_n^* = \epsilon_n^{0*}$. The breakdown point of the proposed AGTE-STLS and any other estimator would be smaller than ϵ_n^* if there exist $m < n\epsilon_n^*$, an index set I_m of size m, and sequences of points $\{\tilde{y}_i^s, \tilde{x}_i^s\}_{s\in\mathbb{N}}, i \in I_m$, such that the respective estimates $\hat{\beta}_n^s$ at samples $C_m^s = \{y_i, x_i\}_{i\in\{1,\ldots,n\}\setminus I_m} \cup \{\tilde{y}_i^s, \tilde{x}_i^s\}_{i\in I_m}$ diverge, $\left\|\hat{\beta}_n^s\right\| \to \infty$ as $s \to \infty$. Note that such a breakdown of (initial or adaptive) GTE-STLS can however occur only if $\|(\tilde{y}_i^s, \tilde{x}_i^s)\| \to \infty$ as $s \to \infty$ at least for some $i \in I_m$: if $C_m^s \subset U(0, K)$ for all $s \in \mathbb{N}$ and some large $K > 0$, identified (GTE-)STLS estimates are finite since all data points are uniformly bounded (cf. Powell, 1986, the proof of Theorem 2). Additionally, identified (GTE-)STLS estimates are also finite if there is some finite $K > 0$ and $\beta \in \mathbb{R}^p$ such that C_m^s lies for all $s \in \mathbb{N}$ within a band around some regression line, $C_m^s \subset \{(x, y) : |y - \max\{y/2, x^\top\beta\}| < K\}$ for $s \in \mathbb{N}$ (for both claims, see the proof of Theorem 5 in Čížek, 2007a, which requires that the data are in a general position). Hence, the contaminated points can cause breakdown by definition (6) only if the norm of $\tilde{r}_i^s(\hat{\beta}_n^s) = \tilde{y}_i^s - \max\{\tilde{y}_i^s/2, \tilde{x}_i^{s\top}\hat{\beta}_n^s)$ increases above any bound as $s \to \infty$ for a sufficient number of observations $i \in \{1, \ldots, n\}$. Considering now an arbitrary, but fixed sequence $C_m^s, s \in \mathbb{N}$, of samples and the initial estimator and its estimates $\hat{\beta}_n^s$ at C_m^s, let us denote $I_\infty = \{i = 1, \ldots, n : \lim_{s\to\infty} \|\tilde{r}_i^s(\hat{\beta}_n^s)\| \to \infty\}$ and $|I_\infty|$ the number of such points.

Next, $m < n\epsilon_n^*$ implies that the initial estimator does not break down, and therefore, $|I_\infty| \leq m$ because non-contaminated points are fixed and finite and the estimates $\hat{\beta}_n^s$ are uniformly bounded in $s \in \mathbb{N}$ as well. Given that GTE-STLS can trim $n - h_n$ observations from its objective function, it will not break down in contaminated samples C_m^s if $h_n \leq n - |I_\infty|$ ($n - h_n \geq |I_\infty|$ observations with the largest residuals are trimmed from the objective function). To prove that AGTE-STLS does not break down at C_m^s, $s \in \mathbb{N}$, we therefore have to show that the adaptively chosen amount of trimming $h_n^a = n - [d_n n] \leq n - |I_\infty|$. To verify this claim, we have to analyze (11) and its elements. The first observation concerns $G_n^0(t\hat{\sigma}_n^0)$ in (11). Because $m < n\epsilon_n^*$, $\hat{\sigma}_n^0$ does not break down at C_m^s and there is some $\delta > 0$ such that $\delta \leq \hat{\sigma}_n^0 \leq 1/\delta$. For any $t > 0$, $G_n^0(t\hat{\sigma}_n^0) \to 1 - |I_\infty|/n$ as $s \to \infty$ because $|I_\infty|$ residuals increase above any t. The second observation concerns $F_{rn,\alpha}(t)$ defined in (10), which can be bounded from below by $(t \geq 0)$

$$F_{rn,\alpha}(t) \geq \frac{\Phi(t)-\Phi(-t)}{\hat{n}_\alpha} \sum_{i=1}^{n} I(t \leq x_i^\top \hat{\beta}_n^0/\hat{\sigma}_n^0) + \frac{\Phi(t)}{\hat{n}_\alpha} \sum_{i=1}^{n} I(\max\{-t, C_n(\alpha)\} < x_i^\top \hat{\beta}_n^0/\hat{\sigma}_n^0 < t).$$

Using $t_0 = \min\{|C_n(\alpha)|, \Phi^{-1}[1-1/(8n)], c\}$, it follows $F_{rn,\alpha}(t_0) \geq -1/(4n) + 1 - 1/(8n) = 1-3/(8n)$. Consequently, we can find $s_0 \in \mathbb{N}$ such that $F_{rn,\alpha}(t_0) - G_n^0(t_0\hat{\sigma}_n^0) \geq |I_\infty|/n - 3/(8n)$ for all $s \geq s_0$, and from equation (11), $h_n^a = n - [d_n n] \leq n - [|I_\infty| - 3/8] = n - |I_\infty|$. This concludes the proof as we have just showed that the adaptively chosen trimming h_n^a satisfies $h_n^a \leq n - |I_\infty|$, GTE-STLS with adaptive trimming does not break down at $C_m^s, s \in \mathbb{N}$, for $m < n\epsilon_n^*$, and the breakdown point of AGTE-STLS must thus be at least ϵ_n^*. $\square$

Proof of Theorem 3: Without loss of generality, we will prove the result only for the estimator ONE-SCLS; the claim for ONE-STLS can be verified along the same lines because the proof relies on the fact that the contribution of a response y_i to the functions (13) and (16) is capped by $2x_i^\top \hat{\beta}_n^0$ in both cases.

First, consider an arbitrary (possibly contaminated) sample $(x_i, y_i)_{i=1}^n$ and let us denote $C = \min\{0, \min_{i=1,\ldots,n} \min_{j=1,\ldots,p} x_{ij}\}$. Note that the SCLS estimates $\hat{\beta}_n$ and $\tilde{\beta}_n$ computed for data $(x_i, y_i)_{i=1}^n$ and $(\tilde{x}_i^c = x_i + (C, \ldots, C)^\top, y_i)_{i=1}^n$, respectively, differ just by the value of the intercept: $\hat{\beta}_n = \tilde{\beta}_n - \Delta_C$, where $\Delta_C = ((C, \ldots, C)^\top \hat{\beta}_n^{-1}/(C+1), 0, \ldots, 0)^\top$ and $\hat{\beta}_n^{-1}$ denotes the vector of slope coefficient $\hat{\beta}_n$, that is, $\hat{\beta}_n$ without the intercept.

Next, using this observation in the context of equation (13) for model CM, we can write

$$\begin{aligned}
\hat{\beta}_n^{(ONE-SCLS)}(\hat{\beta}_n^0) &= \left[\sum_{i=1}^{n} x_i x_i^\top I(x_i^\top \hat{\beta}_n^0 > 0)\right]^{-1} \sum_{i=1}^{n} x_i \min\{y_i, 2x_i^\top \hat{\beta}_n^0\} I(x_i^\top \hat{\beta}_n^0 > 0) \\
&= -\Delta_C + \left[\sum_{i=1}^{n} \tilde{x}_i^c \tilde{x}_i^{c\top} I(x_i^\top \hat{\beta}_n^0 > 0)\right]^{-1} \sum_{i=1}^{n} \tilde{x}_i^c \min\{y_i, 2\tilde{x}_i^{c\top} \hat{\beta}_n^0\} I(x_i^\top \hat{\beta}_n^0 > 0).
\end{aligned}$$

Because $\tilde{x}_{ij}^c \geq 0$ for all $i, j \in \{1, \ldots, n\}$, the trivial inequality

$$-2\sum_{i=1}^{n} \tilde{x}_i^c \tilde{x}_i^{c\top} \hat{\beta}_n^0 I(x_i^\top \hat{\beta}_n^0 > 0) \leq \sum_{i=1}^{n} \tilde{x}_i^c \min\{y_i, 2\tilde{x}_i^{c\top} \hat{\beta}_n^0\} I(x_i^\top \hat{\beta}_n^0 > 0) \leq 2\sum_{i=1}^{n} \tilde{x}_i^c \tilde{x}_i^{c\top} \hat{\beta}_n^0 I(x_i^\top \hat{\beta}_n^0 > 0)$$

implies that

$$\|\hat{\beta}_n^{(ONE-SCLS)}(\hat{\beta}_n^0) + \Delta_C\| \leq \left\| 2\left[\sum_{i=1}^{n} \tilde{x}_i^c \tilde{x}_i^{c\top} I(x_i^\top \hat{\beta}_n^0 > 0)\right]^{-1} \sum_{i=1}^{n} \tilde{x}_i^c \tilde{x}_i^{c\top} I(x_i^\top \hat{\beta}_n^0 > 0) \hat{\beta}_n^0 \right\| \leq 2\|\hat{\beta}_n^0\|.$$

Since the first component of Δ_C is a finite multiple of $\hat{\beta}_n^{-1}$ for any C and all other components of Δ_C are zero, the norm of the slope coefficients of $\hat{\beta}_n^{(ONE-SCLS)}$ is bounded by $2\|\hat{\beta}_n^0\|$, $\|\Delta_C\| \leq 2\|\hat{\beta}_n^0\|$, and finally, $\|\hat{\beta}_n^{ONE-SCLS}\| \leq 4\|\hat{\beta}_n^0\|$.

Consequently at any sample, where the initial estimator does not break down and has a bounded parameter estimates, the ONE-SCLS estimator does not break down too because its coefficients are finite and bounded by a constant proportional to the norm of the initial estimates. The breakdown points of the initial and one-step SCLS are thus the same. □

Proof of Theorem 4: Let us first derive the asymptotic distribution of GTE-STLS with a fixed trimming $h_n = [\lambda n]$. The asymptotic normality of GTE-STLS follows from Čížek (2007b, Theorem 3.2). We will thus verify its assumptions, which are in the majority of cases directly contained or implied by Assumptions A and the assumptions of the theorem. The exceptions, which needs to be verified, are mainly model-specific assumptions F2 (the objective function and normal equations have to form VC classes of functions), I2 (identification of the model parameters), I3 (the Fisher consistency of the estimator), and assumption D3 of Čížek (2007b). The last one, assumption D3, is under Assumptions A verified in Čížek (2006, Lemma 2), so we concentrate on F2, I2, and I3 now.

(F2) We have to prove that $F_0 = \{s(x_i, y_i; \beta) | \beta \in U(\beta^0, \delta)\}$ and $F_1 = \{s'(x_i, y_i; \beta) | \beta \in U(\beta^0, \delta)\}$ form VC classes, where $s(x_i, y_i; \beta) = (y_i - \max\{y_i/2, x_i^\top \beta\})^2 = \min\{y_i/2, y_i - x_i^\top \beta\}^2$ and $s'(x_i, y_i; \beta) = -2x_i(y_i - x_i^\top \beta) I(y_i < 2x_i^\top \beta)$. Since $\{x_i^\top \beta | \beta \in B\}$ is a finite-dimensional vector space, y_i is a constant independent of β, and z^2 is a monotonic function for $z > 0$, F_0 is a VC class of functions (van der Vaart and Wellner, 1996, Lemmas 2.6.15 and 2.6.18). Next, observe that $s'(x_i, y_i; \beta) = -2x_i(y_i - x_i^\top \beta) I(y_i - x_i^\top \beta < y_i/2)$ and $C_i = y_i/2$ and x_i are constants independent of β. Since $\{t_i = y_i - x_i^\top \beta | \beta \in B\}$ is a finite-dimensional vector space, $t_i I(t_i < C_i) = \min\{t_i, C_i I(t_i < C_i)\}$, where both function t_i and $C_i I(t_i < C_i)$ are monotonic, F_1 is also a VC class of functions (van der Vaart and Wellner, 1996, Lemmas 2.6.15 and 2.6.18).

(I2) The identification of the STLS estimates under Assumptions A was proved by Powell (1986). To prove that GTE-STLS is also identified, we therefore have to show that $s(x_i, y_i; \beta) = \min\{y_i/2, y_i - x_i^\top \beta\}^2$ (first-order) stochastically dominates $s(x_i, y_i; \beta^0)$ (see Čížek, 2007b). This is equivalent to proving that $\sqrt{s(x_i, y_i; \beta)} = |\min\{y_i/2, y_i - x_i^\top \beta\}|$ first-order stochastically dominates $\sqrt{s(x_i, y_i; \beta^0)}$. Denoting the unconditional distribution function of $\sqrt{s(x_i, y_i; \beta)}$ by F_r^{β}, we thus have to verify $F_r^{\beta}(t) \leq F_r^{\beta^0}(t)$ for all $t \geq 0$.

To achieve this, we first compute the distribution function $F_{r|\xi,\zeta}^{\beta}$ of the absolute value of standardized symmetrically-trimmed residuals $|r_{is}(\beta)| = |\min\{y_i/2, y_i - x_i^\top \beta\}| = |\min\{(\varepsilon_i + x_i^\top \beta^0)/2, \varepsilon_i + x_i^\top \beta^0 - x_i^\top \beta\}| = |\min\{(\varepsilon_i + \xi_i)/2, \varepsilon_i + \xi_i - \zeta_i\}|$ conditional on $\xi_i = x_i^\top \beta^0$ and $\zeta_i = x_i^\top \beta$. Since $\varepsilon_i + \xi_i > 0$ in truncated samples, $|r_{is}(\beta)| \leq t$ holds if $0 \leq (\varepsilon_i + \xi_i)/2 \leq t$ and $\varepsilon_i + \xi_i - \zeta_i \geq -t$ or if $-t \leq \varepsilon_i + \xi_i - \zeta_i \leq t$; in other words, $|r_{is}(\beta)| \leq t$ holds if $-t - \xi_i + \zeta_i \leq \varepsilon_i \leq 2t - \xi_i$ or $-t - \xi_i + \zeta_i \leq \varepsilon_i \leq t - \xi_i + \zeta_i$. Note that the first upper bound $2t - \xi_i$ dominates the latter upper bound $t - \xi_i + \zeta_i$ if $t > \zeta_i$ and the first upper bound $2t - \xi_i$ is greater than or equal to the common lower bound $-t - \xi_i + \zeta_i$ if $t \geq \zeta_i/3$. Moreover, the lower bound $-t - \xi_i + \zeta_i$ is greater than the truncation point $-\xi_i$ if $t \leq \zeta_i$. We can thus write (F denotes here the distribution function ε conditional on x as we can implicitly condition on x throughout this step)

$$F_{r|\xi,\zeta}^{\beta}(t|\xi_i = \xi, \zeta_i = \zeta) = \begin{cases} [F(t - \xi + \zeta) - F(-t - \xi + \zeta)]/[1 - F(-\xi)] & \text{if } 0 \leq t \leq \zeta, \\ [F(2t - \xi) - F(-\xi)]/[1 - F(-\xi)] & \text{if } \zeta < t. \end{cases} \tag{19}$$

Since the unconditional distribution function F_r^{β} of $|r_{is}(\beta)|$ can be then expressed as $F_r^{\beta}(t) = \int F_{r|\xi,\zeta}^{\beta}(t|\xi = x^\top \beta^0, \zeta = x^\top \beta) dF_x(x)$, where F_x denotes the cumulative distribution function of explanatory variables, we only have to prove for any $x \in \mathbb{R}^p$ and $\beta \in B$ that for $t \geq 0$

$$F_{r|\xi,\zeta}^{\beta}(t|x^\top \beta^0, x^\top \beta) \leq F_{r|\xi,\zeta}^{\beta^0}(t|x^\top \beta^0, x^\top \beta^0).$$

As F is assumed to be unimodal and symmetric, $F(t + a) - F(-t + a)$ is non-increasing in a for $a \geq 0$ and non-decreasing in a for $a < 0$: the first derivative of $F(t + a) - F(-t + a)$ equals $f(t + a) - f(-t + a)$, which is non-positive for $a > 0$ and non-negative for $a < 0$. It follows that

$$\frac{F(t - x^\top \beta^0 + x^\top \beta) - F(-t - x^\top \beta^0 + x^\top \beta)}{1 - F(-x^\top \beta^0)} \leq \frac{F(t) - F(-t)}{1 - F(-x^\top \beta^0)}$$

if $0 \leq t \leq \min\{x^\top \beta^0, x^\top \beta\}$ ($a = -x^\top \beta^0 + x^\top \beta$); that

$$\frac{F(t - x^\top \beta^0 + x^\top \beta) - F(-t - x^\top \beta^0 + x^\top \beta)}{1 - F(-x^\top \beta^0)} \leq \frac{F(t - x^\top \beta^0 + t) - F(-t - x^\top \beta^0 + t)}{1 - F(-x^\top \beta^0)}$$

if $\xi = x^\top\beta^0 \le x^\top\beta = \zeta$ and $\xi < t \le \zeta$ ($a_1 = -x^\top\beta^0 + x^\top\beta \ge -x^\top\beta^0 + t = a_2 > 0$); that

$$\frac{F(2t - x^\top\beta^0) - F(-x^\top\beta^0)}{1 - F(-x^\top\beta^0)} \le \frac{F(t) - F(-t)}{1 - F(-x^\top\beta^0)}$$

if $\zeta = x^\top\beta < x^\top\beta^0 = \xi$ and $\zeta \le t < \xi$ ($a = t - x^\top\beta^0 < 0$); and that

$$\frac{F(2t - x^\top\beta^0) - F(-x^\top\beta^0)}{1 - F(-x^\top\beta^0)}$$

is independent of β. Hence, we obtain $F^{\beta}_{r|\xi,\zeta}(t|x^\top\beta^0, x^\top\beta) \le F^{\beta^0}_{r|\xi,\zeta}(t|x^\top\beta^0, x^\top\beta^0)$ from (19) for any $x \in \mathbb{R}^p$ and $t \ge 0$. Thus, $F^{\beta}_r(t) \le F^{\beta^0}_r(t)$ for all $t \ge 0$ and the identification of GTE-STLS is verified. Moreover for any $\beta \ne \beta^0$, the strict unimodality of F around 0 (Assumption A2) and the continuity of $F^{\beta}_r(t)$ in β implies that the inequality is sharp, $F^{\beta}_r(t) < F^{\beta^0}_r(t)$, if there is $x \in \mathbb{R}^p$ from the support of the distribution function of x such that $x^\top\beta \ne x^\top\beta^0$ and $x^\top\beta^0 > 0$. This is guaranteed by the full-rank condition in Assumption A1.

(I3) To verify the Fisher consistency of GTE-STLS, we have to show that $\mathsf{E}\{s'(x_i, y_i; \beta^0)| s(x_i, y_i; \beta^0) \le s_{[h_n]}(\beta^0)\} = 0$, where $s'(x_i, y_i; \beta^0) = -2x_i(y_i - x_i^\top\beta^0)I(y_i < 2x_i^\top\beta^0)$. This follows directly from Assumption A2 because, conditionally on x_i, ε_i is truncated at $-x_i^\top\beta^0$ (y_i is truncated at 0) and $\varepsilon_i I(|\varepsilon_i| < x_i^\top\beta^0)$ is symmetrically distributed. Hence,

$$x_i\,\mathsf{E}\{\varepsilon_i I(\varepsilon_i < x_i^\top\beta^0)|x_i, \min\{y_i/2, \varepsilon_i\}^2 \le \min\{y_i/2, \varepsilon_i\}^2_{[h_n]}\} = 0.$$

Consequently, Čížek (2007b, Theorem 3.2) implies that the GTE-STLS estimator is asymptotically normal with the variance matrix $V(\lambda)$. This result was derived using the asymptotic linearity of the normal equations of GTE (Čížek, 2007b, Lemma A.7), which can be stated as

$$\sqrt{n}(\hat\beta_n^{h_n} - \beta^0) = n^{-1/2}M_s^{-1}(\lambda)\sum_{i=1}^{n} s'(x_i, y_i; \beta^0) I\left\{s(x_i, y_i; \beta^0) \le s_{[h_n]}(\beta^0)\right\} + o_p(1), \qquad (20)$$

where $\hat\beta_n^{h_n}$ denotes the GTE-STLS using trimming h_n, $s(x_i, y_i; \beta^0) = \min\{\varepsilon_i, y_i/2\}^2$ and $s'(x_i, y_i; \beta^0) = -2x_i\varepsilon_i I(\varepsilon_i < x_i^\top\beta^0)$ (see verification of assumption F2), $s_{[h_n]}(\beta^0)$ denotes the h_nth smallest order statistics of $s(x_i, y_i; \beta)$, and $M_s(\lambda)$ is a non-singular matrix, which depends on the trimming h_n only by means of $\lambda = \lim_{n\to\infty} h_n/n$.

We can now use the expression (20) to prove the second claim of the theorem that the asymptotic distribution does not change if random trimming h_n^a is used: $\sqrt{n}(\hat{\beta}_n^{h_n} - \hat{\beta}_n^{h_n^a}) = o_p(1)$ if $h_n/n \to \lambda$ and $h_n^a/n \to \lambda$ as $n \to \infty$. First, let us note that $s_{[h_n]}(\beta^0)$ and $s_{[h_n^a]}(\beta^0)$ have the same limit Q_λ in probability, the λ-quantile of the $s(x_i, y_i; \beta^0)$ distribution (Čížek, 2007b, Lemma A.3). Equation (20) then implies that

$$\sqrt{n}(\hat{\beta}_n^{h_n} - \hat{\beta}_n^{h_n^a}) = n^{-1/2} M_s^{-1}(\lambda) \sum_{i=1}^{n} s'(x_i, y_i; \beta^0) \Delta I\{s(x_i, y_i; \beta^0), h_n\} + o_p(1) \tag{21}$$

$$- n^{-1/2} M_s^{-1}(\lambda) \sum_{i=1}^{n} s'(x_i, y_i; \beta^0) \Delta I\{s(x_i, y_i; \beta^0), h_n^a\} + o_p(1), \tag{22}$$

where $\Delta I\{s(x_i, y_i; \beta^0), h) = I\{s(x_i, y_i; \beta^0) \leq s_{[h]}(\beta^0)\} - I\{s(x_i, y_i; \beta) \leq Q_\lambda\}$. We will show that both sums (21) and (22) are negligible in probability. To show that

$$n^{-1/2} \sum_{i=1}^{n} s'(x_i, y_i; \beta^0)[I\{s(x_i, y_i; \beta^0) \leq s_{[h]}(\beta^0)\} - I\{s(x_i, y_i; \beta) \leq Q_\lambda\}] \tag{23}$$

is negligible in probability (h represents h_n or h_n^a), note that

$$\mathsf{E} \left| n^{1/4} s'(x_i, y_i; \beta^0) \left[I\{s(x_i, y_i; \beta^0) \leq s_{[h]}(\beta^0)\} - I\{s(x_i, y_i; \beta) \leq Q_\lambda\} \right] \right|^l = \mathcal{O}(1)$$

for $l = 1, 2$ by Assumption A1 and Čížek (2007b, Corollary A.5) and the summands in (23) form a stationary sequence of random variables with zero means and finite variances. Thus, the law of large numbers for mixingales (e.g., Davidson, 1994, Corollary 20.16) leads to

$$n^{-3/4} \sum_{i=1}^{n} n^{1/4} s'(x_i, y_i; \beta^0) \left[I\{s(x_i, y_i; \beta^0) \leq s_{[h]}(\beta^0)\} - I\{s(x_i, y_i; \beta) \leq Q_\lambda\} \right] \to 0$$

as $n \to \infty$. This proves that both sums in (21) and (22) are negligible in probability as $n \to \infty$ and concludes the proof. □

Proof of Theorem 5: Without loss of generality, let us derive this result only in the case of ONE-STLS; the proof for ONE-SCLS can follow the same steps.

First, let us recall that Powell (1986, Theorem 2) derived the asymptotic normality of an estimator (STLS) defined by the equation

$$n^{-1/2}\sum_{i=1}^{n}\psi(\varepsilon_i, x_i; \hat{\beta}_n) = o_p(1), \tag{24}$$

where $\psi(\varepsilon_i, x_i; \beta) = x_i(\varepsilon_i + x_i^\top\beta^0 - x_i^\top\beta)I(\varepsilon_i + x_i^\top\beta^0 - x_i^\top\beta < x_i^\top\beta)$, provided that $\hat{\beta}_n$ is consistent. Thus, any consistent estimator that satisfies the normal equations (24) has the same asymptotic distribution as STLS, and by equation (A.22) in Powell (1986, Theorem 3), it satisfies

$$\sqrt{n}(\hat{\beta}_n - \beta^0) = n^{-1/2}C_n^{-1}\sum_{i=1}^{n}\psi(\varepsilon_i, x_i; \beta^0) + o_p(1), \tag{25}$$

where C_n is a positive definite matrix independent of β. The claim of the theorem then follows from the fact that the right hand side of (25) is independent of $\hat{\beta}_n$.

To prove that $\hat{\beta}_n^{(ONE-STLS)}$ satisfies (24), we need the following lemma.

Lemma 1 *Let x_i and γ_n be absolutely integrable random variables, z_i have finite second moments, and $\hat{\beta}_n^0$ be an n^α-consistent estimator of β^0, $n^\alpha(\hat{\beta}_n^0 - \beta^0) = \mathcal{O}_p(1)$ as $n \to \infty$. If $f(x_i, \gamma_n; \hat{\beta}_n^0)$ is linear in all its arguments, Assumption A imply*

$$\mathsf{E}[I\{\varepsilon_i < f(x_i, \gamma_n; \hat{\beta}_n^0)\} - I\{\varepsilon_i < f(x_i, \gamma_n; \beta^0)\}] = O(n^{-\alpha}). \tag{26}$$

Further, if $(z_i, x_i, \varepsilon_i)_{i=1}^n$ forms a stationary sequence of random vectors,

$$\frac{1}{n}\sum_{i=1}^{n} z_i[I\{\varepsilon_i < f(x_i, \gamma_n; \hat{\beta}_n^0)\} - I\{\varepsilon_i < f(x_i, \gamma_n; \beta^0)\}] = o_p\left(n^{-\alpha+\delta}\right)$$

as $n \to \infty$ for any $\delta > 0$.

Proof: Let us denote $\Delta I(\varepsilon_i, x_i, \gamma_n; \beta) = I\{\varepsilon_i < f(x_i, \gamma_n; \beta)\} - I\{\varepsilon_i < f(x_i, \gamma_n; \beta^0)\}$. We first consider a neighborhood $U(\beta^0, n^{-\alpha}M)$ for some $M > 0$ and suprema over all $\beta \in U(\beta^0, n^{-\alpha}M)$. Conditioning on x_i and γ_n and denoting $K > 0$ the uniform upper bound on the (conditional) density of ε_i (Assumption A2),

$$P[\exists\beta{:}\Delta I(\varepsilon_i, x_i, \gamma_n; \beta) \neq 0|x_i, \gamma_n] \leq K\sup_{\beta}|f(x_i, \gamma_n; \beta) - f(x_i, \gamma_n; \beta^0)| \leq K|\tilde{f}(x_i, \gamma_n)|\sup_{\beta}\|\beta - \beta^0\|,$$

where $\tilde{f}$ is a linear function of its arguments and the second inequality follows from the linearity of f. Hence,

$$\mathsf{E}\sup_{\beta} |\Delta I(\varepsilon_i, x_i, \gamma_n; \beta)| \leq K\,\mathsf{E}\,|\tilde{f}(x_i, \gamma_n)| \cdot n^{-\alpha} M. \tag{27}$$

The n^{α}-consistency of $\hat{\beta}_n^0$ implies that for any $\varepsilon > 0$ there is M such that $\hat{\beta}_n^0 \in U(\beta^0, n^{-\alpha}M)$ with probability larger than $1-\varepsilon$. Since the upper bound in (27) depends on $\hat{\beta}_n^0$ only via M, (27) implies (26) for $\varepsilon \to 0$.

Next, let us consider $\sum_{i=1}^n z_i \Delta I(\varepsilon_i, x_i, \gamma_n; \hat{\beta}_n^0)/n$. By Markov inequality, $P(X > K) \leq \mathsf{E}\,X/K$ for non-negative random variable X, it holds for $\delta > 0$

$$P\left[\left|\sum_{i=1}^n z_i \Delta I(\varepsilon_i, x_i, \gamma_n; \hat{\beta}_n^0)/n\right| > K n^{-\alpha+\delta}\right] \leq K^{-1} n^{\alpha-\delta}\,\mathsf{E}\,|z_i \Delta I(\varepsilon_i, x_i, \gamma_n; \hat{\beta}_n^0)|.$$

Further, by the Chebyshev inequality and the claim (26) imply for $n \to \infty$

$$n^{\alpha-\delta}\,\mathsf{E}\,|z_i \Delta I(\varepsilon_i, x_i, \gamma_n; \hat{\beta}_n^0)| \leq \mathsf{E}\,z_i^2 \cdot n^{\alpha-\delta}\,\mathsf{E}\,|\Delta I(\varepsilon_i, x_i, \gamma_n; \hat{\beta}_n^0)| = o(1),$$

which confirms the second claim that $\sum_{i=1}^n z_i \Delta I(\varepsilon_i, x_i, \gamma_n; \hat{\beta}_n^0)/n = o_p\big(n^{-\alpha+\delta}\big)$ as $n \to \infty$. $\square$

Let us now verify that $\hat{\beta}_n^{(ONE-STLS)}$ is consistent and satisfies (24). First, recall that $\sqrt{n}(\hat{\beta}_n^0 - \beta^0) = \mathcal{O}_p(1)$ as $n \to \infty$ and that equation (16) states

$$\hat{\beta}_n^T(\hat{\beta}_n^0) = \frac{Q_n^{-1}}{n} \sum_{i=1}^n x_i y_i I(y_i < 2x_i^\top \hat{\beta}_n^0), \tag{28}$$

where $Q_n = \sum_{i=1}^n x_i x_i^\top I(y_i < 2x_i^\top \hat{\beta}_n^0)/n$. This implies

$$\begin{aligned}
\hat{\beta}_n^T(\hat{\beta}_n^0) - \beta^0 &= \frac{Q_n^{-1}}{n} \sum_{i=1}^n x_i \varepsilon_i \{I(\varepsilon_i < 2x_i^\top \hat{\beta}_n^0 - x_i^\top \beta^0) - I(\varepsilon_i < x_i^\top \beta^0)\} && (29)\\
&+ \frac{Q_n^{-1}}{n} \sum_{i=1}^n x_i \varepsilon_i I(\varepsilon_i < x_i^\top \beta^0). && (30)
\end{aligned}$$

By Assumption A1 and the law of large numbers (e.g., Davidson, 1994, Corollary 20.16), Q_n converges to a positive definite matrix. Hence to prove consistency of $\hat{\beta}_n^T(\hat{\beta}_n^0)$, we only have to show that the averages (29) and (30) are negligible in probability as $n \to \infty$. For average (29), this follows directly from Lemma 1. For average (30), note that the expectations of its summands equal zero, $\mathsf{E}\{x_i \varepsilon_i I(\varepsilon_i < x_i^\top \beta^0)\} = 0$, and their variances are finite by Assumptions

A1 and A2. The law of large numbers for martingale differences (e.g., Davidson, 1994, Theorem 20.10) thus implies that (30) converges to zero in probability. Hence, estimator $\hat{\beta}_n^{(ONE-STLS)} = \hat{\beta}_n^T(\hat{\beta}_n^0)$ is consistent. Since $\hat{\beta}_n^0$ is $\sqrt{n}$-consistent, the rate of convergence of $\hat{\beta}_n^{(ONE-STLS)}$ is at least $n^{1/2-\delta}$ for some $\delta > 0$ by Lemma 1 for (29) and by Davidson (1994, Theorem 20.10) for (30).

Next, we verify that $\hat{\beta}_n^{(ONE-STLS)}$ satisfies (24). Similarly to (29)–(30), (28) also implies (by subtracting $\hat{\beta}_n^T(\hat{\beta}_n^0)$ from both sides and substituting $y_i = x_i^\top \beta^0 + \varepsilon_i$) that

$$\begin{aligned}
0 &= \frac{1}{n}\sum_{i=1}^{n} x_i\{\varepsilon_i + x_i^\top\beta^0 - x_i^\top\hat{\beta}_n^T(\hat{\beta}_n^0)\}I(\varepsilon_i + x_i^\top\beta^0 - x_i^\top\hat{\beta}_n^T(\hat{\beta}_n^0) < 2x_i^\top\hat{\beta}_n^0 - x_i^\top\hat{\beta}_n^T(\hat{\beta}_n^0)) && (31)\\
&= \frac{1}{n}\sum_{i=1}^{n} x_i\{\varepsilon_i + x_i^\top\beta^0 - x_i^\top\hat{\beta}_n^T(\hat{\beta}_n^0)\} && (32)\\
&\quad \times\{I(\varepsilon_i + x_i^\top\beta^0 - x_i^\top\hat{\beta}_n^T(\hat{\beta}_n^0) < 2x_i^\top\hat{\beta}_n^0 - x_i^\top\hat{\beta}_n^T(\hat{\beta}_n^0)) - I(\varepsilon_i + x_i^\top\beta^0 - x_i^\top\hat{\beta}_n^T(\hat{\beta}_n^0) < x_i^\top\hat{\beta}_n^T(\hat{\beta}_n^0))\}\\
&+ \frac{1}{n}\sum_{i=1}^{n} x_i\{\varepsilon_i + x_i^\top\beta^0 - x_i^\top\hat{\beta}_n^T(\hat{\beta}_n^0)\}I(\varepsilon_i + x_i^\top\beta^0 - x_i^\top\hat{\beta}_n^T(\hat{\beta}_n^0) < x_i^\top\hat{\beta}_n^T(\hat{\beta}_n^0)). && (33)
\end{aligned}$$

To verify that $\hat{\beta}_n^T(\hat{\beta}_n^0)$ satisfies (24), we thus have show that (32) is negligible in probability as $n \to \infty$ since (33) is identical with (24). The term (32) is however negligible in probability by Lemma 1. This concludes the proof as we have shown that (31)–(33) implies

$$o_p(1) = \frac{1}{n}\sum_{i=1}^{n} x_i\{\varepsilon_i + x_i^\top\beta^0 - x_i^\top\hat{\beta}_n^T(\hat{\beta}_n^0)\}I\{\varepsilon_i + x_i^\top\beta^0 - x_i^\top\hat{\beta}_n^T(\hat{\beta}_n^0) < x_i^\top\hat{\beta}_n^T(\hat{\beta}_n^0)\}$$

and that $\hat{\beta}_n^{(ONE-STLS)} = \hat{\beta}_n^T(\hat{\beta}_n^0)$ thus satisfies (24). □

References

[1] Arabmazar, A. and P. Schmidt, 1982, An investigation of the robustness of the Tobit estimator to non-normality, *Econometrica* **50(4)**, 1055–1063.

[2] Bramati, M. and C. Croux, 2007, Robust estimators for the fixed effects panel data model, *The econometrics journal* **10(3)**, 521–540.

[3] Čížek, P., 2007a, Efficient robust estimation of regression models, CentER Discussion Paper 87/2007, Tilburg University.

[4] Čížek, P., 2007b, General trimmed estimation: robust approach to nonlinear and limited dependent variable models. CentER Discussion Paper 65/2007, Tilburg University; to appear in Econometric Theory.

[5] Davidson, J., 1994, *Stochastic Limit Theory*, Oxford University Press, New York.

[6] Debruyne, M., M. Hubert, S. Portnoy and K. Van den Branden, 2008, Censored depth quantiles, *Computational Statistics & Data Analysis* **52(3)**, 1604–1614.

[7] Gallant, A. R. and D. W. Nychka, 1987, Semi-nonparametric maximum likelihood estimation, *Econometrica* **55(2)**, 363–390.

[8] Genton, M. G. and A. Lucas, 2003, Comprehensive definitions of breakdown points for independent and dependent observations, *Journal of the Royal Statistical Society, Series B* **65**, 81–94.

[9] Gerfin, M., 1996, Parametric and semi-parametric estimation of the binary response model of labour market participation, *Journal of Applied Econometrics* **11(3)**, 321–339.

[10] Gervini, D. and V. J. Yohai, 2002, A class of robust and fully efficient regression estimators, *The Annals of Statistics* **30(2)**, 583–616.

[11] Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw and W. A. Stahel, 1986, *Robust statistics, the approach based on influence function*, Wiley, New York.

[12] Honore, B. E., 1992, Trimmed LAD and least squares estimation of truncated and censored regression models with fixed effects, *Econometrica* **60(3)**, 533–565.

[13] Honore, B. E. and L. Hu, 2007, Estimation of cross sectional and panel data censored regression models with endogeneity, *Journal of Econometrics* **122(2)**, 293–316.

[14] Honore, B., S. Khan and J. L. Powell, 2002, Quantile regression under random censoring, *Journal of Econometrics* **109**, 67–105.

[15] Honore, B. E. and J. L. Powell, 1994, Pairwise difference estimators for censored and truncated regression models, *Journal of Econometrics* **64(l-2)**, 241–278.

[16] Hu, L., 2002, Estimation of a censored dynamic panel data, *Econometrica* **70(6)**, 2499–2517.

[17] Ichimura, H., 1993, Semiparametric least squares (SLS) and weighted SLS estimation of single-index models, *Journal of Econometrics* **58**, 71–120.

[18] Khan, S. and J. L. Powell, 2001, Two-step estimation of semiparametric censored regression models, *Journal of Econometrics* **103**, 73–110.

[19] Krasker, W. S. and R. E. Welsch, 1985, Resistant estimation for simultaneous-equations models using weighted instrumental variables, *Econometrica* **53**, 1475–1488.

[20] Krishnakumar, J. and E. Ronchetti, 1997, Robust estimators for simultaneous equations models, *Journal of Econometrics* **78**, 295–314.

[21] Lee, M. J., 1993, Quadratic mode regression, *Journal of Econometrics* **57**, 1–19.

[22] Lewbel, A., 1998, Semiparametric latent variable model estimation with endogenous or mismeasured regressors, *Econometrica* **66(1)**, 105–121.

[23] Lewbel, A. and O. Linton, 2002, Nonparametric censored and truncated regression, *Econometrica* **70(2)**, 765–779.

[24] Marrona, R. A., O. H. Bustos and V. J. Yohai, 1979, Bias- and efficiency-robustness of general M-estimators for regression with random carriers, in: T. Gasser and M. Rossenblatt (Eds.), *Smoothing techniques for curve estimation*, Lecture Notes in Mathematics 757, Springer, Berlin, pp. 91–116.

[25] Marazzi, A. and V. J. Yohai, 2004, Adaptively truncated maximum likelihood regression with asymmetric errors, *Journal of Statistical Planning and Inference* **122**, 271–291.

[26] Ortelli, C. and F. Trojani, 2005, Robust efficient method of moments, *Journal of Econometrics* **128**, 69–97.

[27] Peracchi, F., 1990, Bounded-influence estimators for the Tobit model, *Journal of Econometrics* **44**, 107–126.

[28] Portnoy, S., 2003, Censored regression quantiles, *Journal of the American Statistical Association* **98**, 1001–1012.

[29] Powell, J. L., 1984, Least absolute deviations estimator for the censored regression model, *Journal of Econometrics* **25**, 303–325.

[30] Powell, J. L., 1986, Symmetrically trimmed least squares estimation for Tobit models, *Econometrica* **54**, 1435–1460.

[31] Ren, J.-J. and M. Gu, 1997, Regression M-estimators with doubly censored data, *The Annals of Statistics* **25(6)**, 2638–2664.

[32] Ren, J.-J., 2003, Regression M-estimators with non-i.i.d. doubly censored data, *The Annals of Statistics* **31(4)**, 1186–1219.

[33] Ronchetti, E. and F. Trojani, 2001, Robust inference with GMM estimators, *Journal of Econometrics* **101**, 37–69.

[34] Rousseeuw, P. J., 1985, Multivariate estimation with high breakdown point, in: W. Grossman, G. Pflug, I. Vincze and W. Wertz (Eds.), *Mathematical statistics and applications, Vol. B*, Reidel, Dordrecht, Netherlands, pp. 283–297.

[35] Rousseeuw, P. J., 1997, Introduction to positive-breakdown methods, in: G. S. Maddala and C. R. Rao (Eds.), *Handbook of statistics, Vol. 15: Robust inference*, Elsevier, Amsterdam, pp. 101–121.

[36] Rousseeuw, P. J. and M. Hubert, 1999, Regression depth, *Journal of the American Statistical Association* **94**, 388–402.

[37] Rousseeuw, P. J. and A. M. Leroy, 2003, *Robust regression and outlier detection*, Wiley, New York.

[38] Salibian-Barrera, M. and R. H. Zamar, 2002, Bootstrapping robust estimates of regression, *The Annals of Statistics* **30(2)**, 556–582.

[39] Sakata, S. and H. White, 1998, High breakdown point conditional dispersion estimation with application to S&P 500 daily returns volatility, *Econometrica* **66(3)**, 529–567.

[40] Santos Silva, J. M. C., 2001, Influence diagnostics and estimation algorithms for Powell's SCLS, *Journal of Business & Economic Statistics* **19(1)**, 55–62.

[41] Simpson, D. G., D. Ruppert and R. J. Carroll, 1992, On one-step GM estimates and stability of inferences in linear regression, *Journal of the American Statistical Association* **87**, 439–450.

[42] Stromberg, A. J. and D. Ruppert, 1992, Breakdown in nonlinear regression, *Journal of American Statistical Association* **87**, 991–997.

[43] Van der Vaart, A. W. and J. A. Wellner, 1996, *Weak convergence and empirical processes: with applications to statistics*, Springer, New York.

[44] Welsh, A. H. and E. Ronchetti, 2002, A journey in single steps: robust one-step M-estimation in linear regression, *Journal of Statistical Planning and Inference* **103**, 287–310.

[45] Willems, G. and S. van Aelst, 2005, Fast and robust bootstrap for LTS, *Computational Statistics & Data Analysis* **48(4)**, 703–715.

[46] Zhou, M., 1992, M-estimation in censored linear models, *Biometrika* **79(4)**, 837–841.